Niveau(x) d'enseignement: 7 Numéro EDUQ : 8570263
Centre d'intérêt: 841

Apprentissage des probabilités et de la statistique selon la méthode B.B.G.R. / Gaétan Beaudoin, Joseph-Marie Briand, Gérald Gagnon, Alain Roy ; [pour le] Cégep de Rimouski. -- Rimouski : Le Collège, c1982.
660 p. en pag. multiple : ill., graph. ; 28 cm.

Bibliogr. à la fin du v.

Descripteur(s): *Mathématiques *Statistique *Méthode d'enseignement *Activité d'apprentissage *Pratique pédagogique *Manuel *Collégial (Niveau)

Identificateur(s) de nom: *Piaget, Jean *Dienes, Z. P

Identificateur(s) de sujet: *Probabilités

Le document présente la méthode BBGR d'apprentissage des probabilités et de la statistique. Il est destiné aux étudiants qui abordent la statistique pour la première fois, en particulier au niveau collégial. La méthode BBGR (d'après le nom des auteurs) veut assurer la compréhension des concepts de base, de leur enchaînement et du symbolisme qui les véhicule, assurer la maîtrise des techniques et développer des habiletés à l'analyse des données, à l'interprétation des résultats et à leur expression. Pour y arriver, les auteurs ont élaboré une démarche qui repose sur le processus de développerment de l'intelligence tel que décrit par Piaget et repris par Z. P. Diénès pour l'enseignement des mathématiques. Elle respecte les étapes d'apprentissage suivantes: intuition, abstraction, symbolisation, généralisation, formalisation.

Pour opérationnaliser les étapes d'apprentissage et la stratégie pédagogique, les auteurs ont créé et présentent dans ce manuel: 1. trois expériences préalables qui visent à faire prendre conscience à l'étudiant qu'il possède déjà certains outils propres aux probabilités et à la statistique mais que ces outils ont leur limite; 2. un laboratoire expérimental divisé en dix expériences, chacune portant sur des concepts importants et formant un tout cohérent; 3. dix textes de symbolisation qui servent à faire un retour, une réflexion sur chaque expérience; 4. un coffre où sont réunis tous les outils vus au laboratoire expérimental et les grandes lignes de la méthode statistique; 5. un laboratoire où sont présentés des problèmes rencontrés dans les différents champs d'application de la statistique; 6. un texte de formalisation qui synthétise et exprime d'une façon concise l'ensemble des concepts déjà vus.

Type de document: d
Financement: Bénéficiaire - Collège de Rimouski
Organisme(s) source(s) - MEQ-DGEC
Programme(s) de subvention - PROSIP

EDUQ 8570263

cégep
rimouski

19.25

Apprentissage des probabilités et de la statistique selon la méthode B.B.G.R.

Gaétan Beaudoin
Joseph-Marie Briand
Gérald Gagnon
Alain Roy

cégep
rimouski

19.25

Apprentissage des probabilités et de la statistique selon la méthode B.B.G.R.

Gaétan Beaudoin
Joseph-Marie Briand
Gérald Gagnon
Alain Roy

AUTEURS:

Gaétan Beaudoin
Joseph-Marie Briand
Gérald Gagnon
Alain Roy

COLLABORATEURS:

Dactylographie: *Ginette D'Amours*

Graphisme: *Robert Casavant*

Conseiller pédagogique: *Jean-Claude Marquis*

Évaluation: *Carmen Allard*
Christian Bernatchez
Réjean Gignac
Claude Parent
Richard Proulx
Céline St-Pierre
Réal Savard
Bertrand Voyer

ORGANISMES:

Ministère de l'éducation du Québec: *D.G.E.C. (par le programme P.E.D.A.G.I.P.)*

Cégep de Rimouski:

Direction des services pédagogiques.

Service des moyens d'enseignement.

Service de l'équipement.

© Cégep de Rimouski 1982

PREFACE

Ce texte constitue le support écrit de la méthode B.B.G.R. d'apprentissage des probabilités et de la statistique. C'est le résultat de plusieurs années de recherche et d'expérimentation. Il s'adresse à tous (toutes) les étudiants(es) qui abordent la statistique pour la première fois et en particulier à ceux (celles) de niveau collégial.

La méthode B.B.G.R. est une nouvelle approche dans l'apprentissage des probabilités et de la statistique. Elle veut assurer la compréhension des concepts de base, de leur enchaînement et du symbolisme qui les véhicule, assurer la maîtrise des techniques et développer des habiletés à l'analyse des données, à l'interprétation des résultats et à leur expression. Pour y arriver, nous avons choisi une démarche qui repose sur le processus de développement de l'intelligence tel que décrit par Piaget et repris par Z.P. Diénès pour l'enseignement des mathématiques. Elle respecte les étapes d'apprentissage suivantes: intuition, abstraction, symbolisation, généralisation, formalisation.

L'apprentissage d'un concept se fait selon un cheminement en spirale c'est-à-dire, à différents niveaux: de l'intuition jusqu'à la formalisation. En plus, la stratégie pédagogique retenue consiste à partir du connu de l'étudiant(e) pour en arriver à l'inconnu et à utiliser l'analogie du coffre d'outils

pour faire prendre conscience de la valeur des outils intellectuels.

Pour opérationnaliser les étapes d'apprentissage et la stratégie pédagogique, nous avons créé:

- trois expériences préalables qui visent à faire prendre conscience à l'étudiant(e) qu'il (elle) possède déjà certains outils propres aux probabilités et à la statistique mais que ces outils ont leur limite.

- un laboratoire expérimental divisé en dix expériences, chacune portant sur des concepts importants et formant un tout cohérent. L'utilisation de matériel pédagogique, principalement des "bonshommes Lego" permet à l'étudiant(e) de visualiser ces notions.

- dix textes de symbolisation qui servent à faire un retour, une réflexion sur chaque expérience. C'est ici que les définitions et les symboles sont donnés.

- un coffre où sont réunis tous les outils vus au laboratoire expérimental et les grandes lignes de la méthode statistique.

- un laboratoire d'applications où sont présentés des problèmes rencontrés dans les différents champs d'application de la statistique. Pour résoudre ces problèmes, l'étudiant(e) doit faire appel aux outils

  et à la méthode acquis au laboratoire expérimental ou rechercher de nouveaux outils présentés dans une sorte de livre de recettes: le manuel d'instructions.

- un texte de formalisation qui synthétise et exprime d'une façon concise et rigoureuse l'ensemble des concepts déjà vus.

Le professeur qui a l'intention d'utiliser la méthode B.B.G.R. aurait avantage à se procurer le guide pédagogique en s'adressant au Cégep de Rimouski.

Nous tenons à remercier toutes les personnes et tous les organismes qui ont rendu possible l'élaboration et l'amélioration de ce volume:

La D.G.E.C. par le biais de son programme P.R.O.S.I.P.;

Le Collège de Rimouski par l'intermédiaire de ses différents services:

- les services pédagogiques et en particulier le service de recherche et d'expérimentation sous la responsabilité de M. Jean-Claude Marquis;

- le service des moyens d'enseignement et le service de l'équipement.

Des remerciements s'adressent aussi aux professeurs suivants pour leurs suggestions.

Mesdames Carmen Allard
Céline St-Pierre

Messieurs Christian Bernatchez
Réjean Gignac
Claude Parent
Richard Proulx
Réal Savard
Bertrand Voyer

tous du Cégep de Rimouski sauf M. Gignac qui est du Cégep de Sainte-Foy.

Enfin, nous sommes reconnaissants envers Madame Ginette D'Amours pour la dactylographie du texte et M. Robert Casavant pour le graphisme.

Les Auteurs

Table des matières

Page

La méthode d'apprentissage des probabilités et de la statistique . . . . . (saumon)

- Schéma

Les expériences préalables . . . . . (bleue)

- Schéma
- Expérience préalable ...

P1 Éveil statistique . . . . . EP 1.1
P2 Coup d'oeil sur ma classe . . . . . EP 2.1
P3 Observation statistique . . . . . EP 3.1

- Du connu vers l'inconnu par les expériences préalables . . . . . SP 1

Le laboratoire expérimental . . . . . (rose)

- Schéma
- Expérience ...

1 Caractéristiques et catégories . . . . . E 1.1
2 Classification . . . . . E 2.1
3 Distribution . . . . . E 3.1
4 Tendance centrale . . . . . E 4.1
5 Dispersion . . . . . E 5.1
6 Corrélation . . . . . E 6.1
7 Population et échantillon . . . . . E 7.1
8 Échantillonnage . . . . . E 8.1
9 Probabilité . . . . . E 9.1
10 Erreur d'échantillonnage . . . . . E 10.1

La symbolisation . . . . . (verte)

- Schéma

Page

- Symbolisation ...

1 Caractéristiques et catégories . . . . . S 1.1
2 Classification . . . . . S 2.1
3 Distribution . . . . . S 3.1
4 Tendance centrale . . . . . S 4.1
5 Dispersion . . . . . S 5.1
6 Corrélation . . . . . S 6.1
7 Population et échantillon . . . . . S 7.1
8 Échantillonnage . . . . . S 8.1
9 Probabilité . . . . . S 9.1
10 Erreur d'échantillonnage . . . . . S 10.1

Le coffre d'outils . . . . . (jaune)

- Coffre d'outils

Le laboratoire d'applications ..... (saumon)

- Schéma
- Application

I. Le rapport du commissaire industriel de la ville de Bécart A 1.1

II. Les activités étudiantes du groupe au cours des dernières vacances estivales A 2.1

III. Étude des plaintes formulées par les clients de la coopérative d'alimentation sur la qualité des pommes de terre A 3.1

IV. Étude du revenu des étudiants(es) en statistique au cours des dernières vacances estivales A 4.1

- Travaux pratiques
  1. Structure temporelle T 1.1
  2. Synthèse T 2.1
  3. Corrélation T 3.1
  4. Probabilité T 4.1

La formalisation ..... (jaune)
- Schéma
  1. Analyse d'un groupe de données F.9
  2. Analyse de l'information recueillie auprès d'une population F.
  3. Analyse de l'information recueillie auprès d'un échantillon F.

Annexes .....
  1. Enquête (verte)
  2. Manuel d'instructions (orange)
     2.1 Les outils statistiques M.
     2.2 La méthode pour étudier les données M.
     2.3 La rédaction d'un rapport M.
  3. Symboles ..... (saumon)

La méthode d'apprentissage des probabilités et de la statistique

MÉTHODE B.B.G.S. D'APPRENTISSAGE DES PROBABILITÉS ET DE LA STATISTIQUE

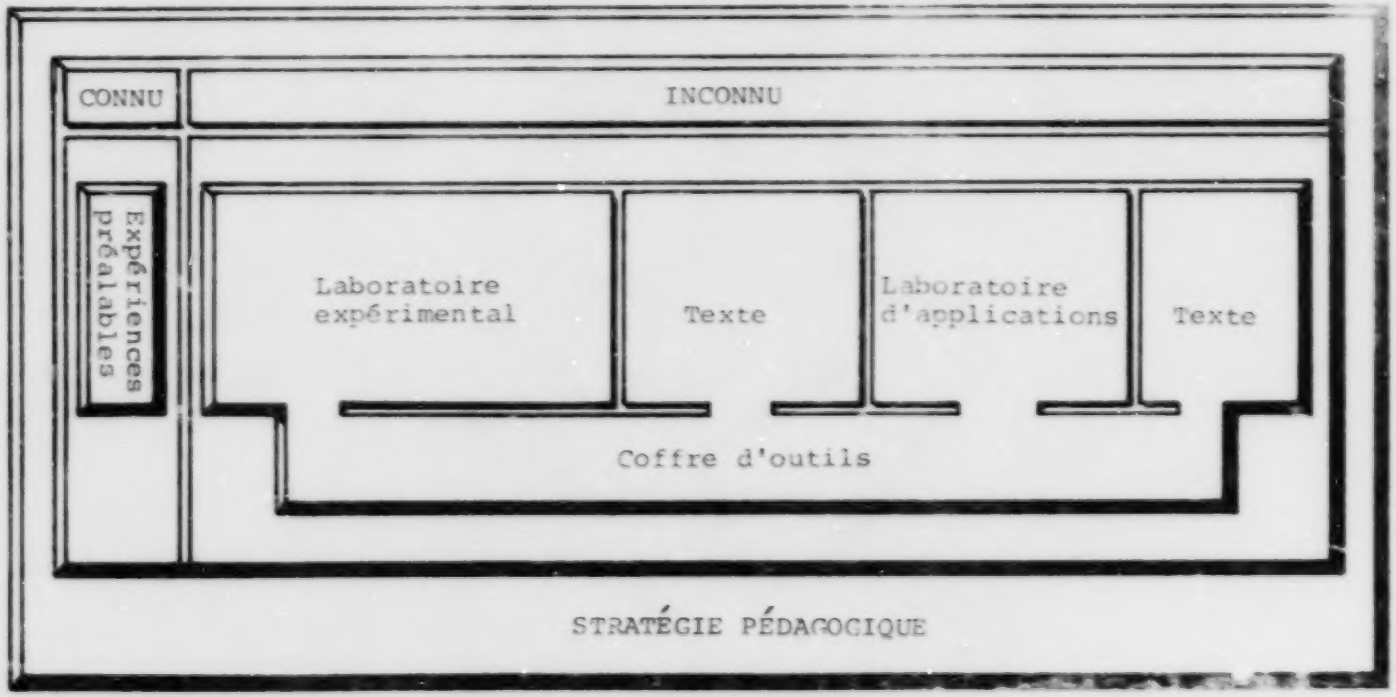
OBJECTIFS:

- Assurer la compréhension des concepts de base, de leur enchaînement, et du symbolisme qui les véhicule.
- Assurer la maîtrise des techniques.
- Développer des habiletés à l'analyse des données, à l'interprétation des résultats et à leur expression.

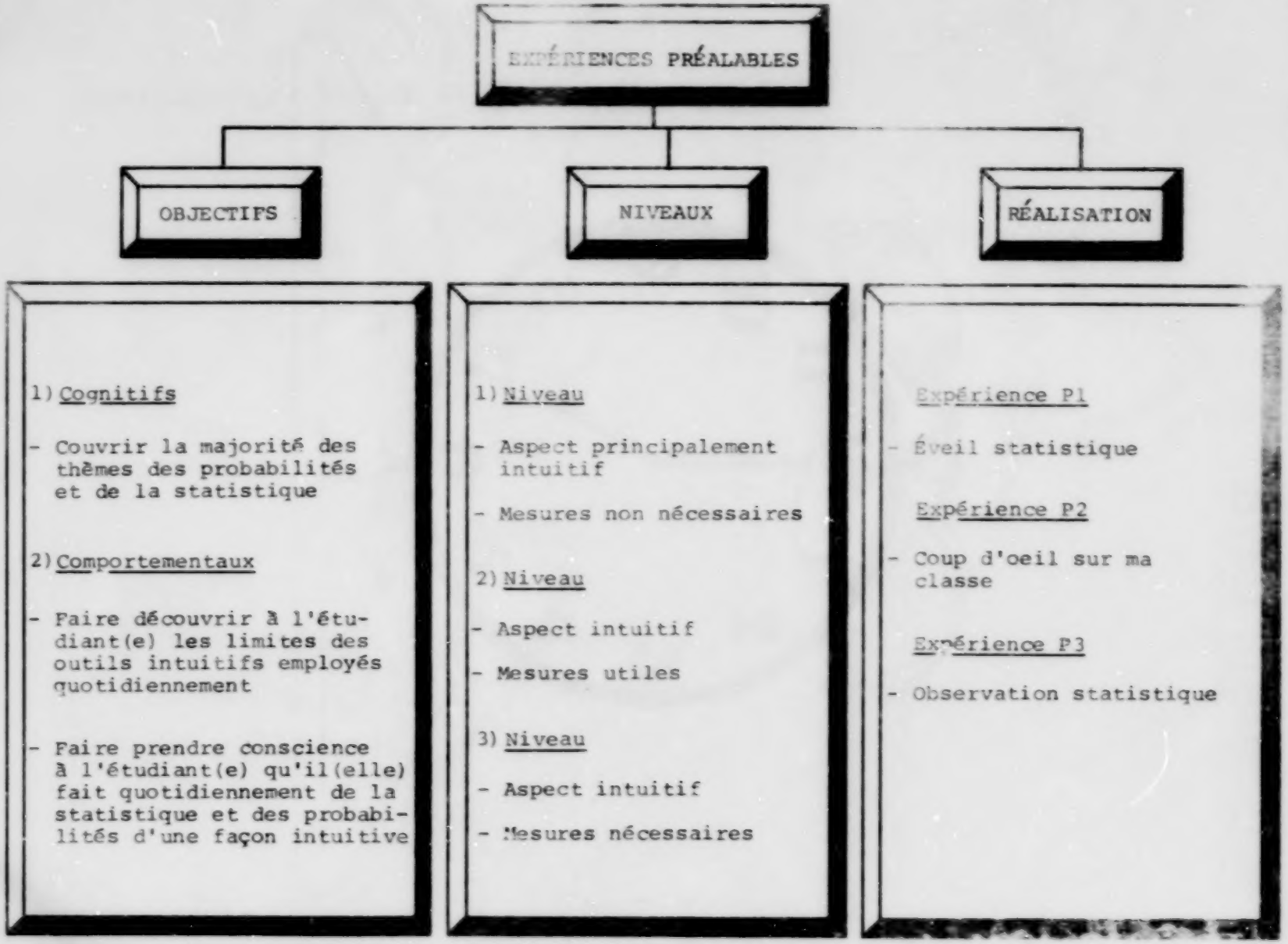
FONDEMENT PÉDAGOGIQUE

ÉTAPES				
Intuition	Abstraction	Symbolisation	Généralisation	Formalisation

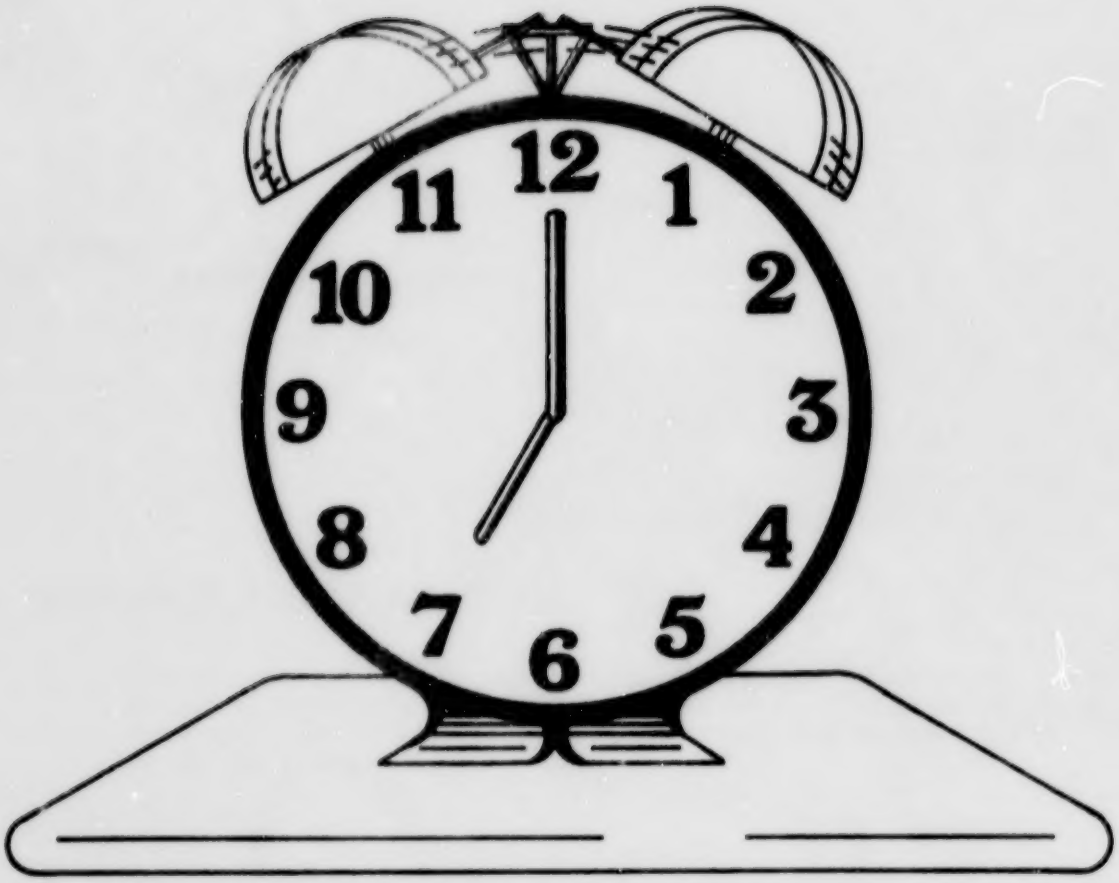
RÉALISATION



Expériences préalables


EXPÉRIENCES PRÉALABLES
OBJECTIFS
NIVEAUX
RÉALISATION
1) Cognitifs
- Couvrir la majorité des thèmes des probabilités et de la statistique
2) Comportementaux
- Faire découvrir à l'étudiant(e) les limites des outils intuitifs employés quotidiennement
- Faire prendre conscience à l'étudiant(e) qu'il(elle) fait quotidiennement de la statistique et des probabilités d'une façon intuitive
1) Niveau
- Aspect principalement intuitif
- Mesures non nécessaires
2) Niveau
- Aspect intuitif
- Mesures utiles
3) Niveau
- Aspect intuitif
- Mesures nécessaires
Expérience P1
- Éveil statistique
Expérience P2
- Coup d'oeil sur ma classe
Expérience P3
- Observation statistique

Éveil statistique



EXPÉRIENCE P 1

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Prendre conscience de notre connaissance intuitive des concepts des probabilités et de la statistique et de notre habilité naturelle à les véhiculer. Également, prendre conscience des limites de la connaissance intuitive.

MATÉRIEL

Un crayon et du papier.

CONTEXTE DE L'EXPÉRIENCE

Cette expérience est une réflexion sur vos heures de lever et coucher. Les questions posées, d'apparence anodine, véhiculent plusieurs concepts des probabilités et de la statistique.

CHEMINEMENT DE L'EXPÉRIENCE

Q 1. a) *À quelle heure t'es-tu levé(e) aujourd'hui?* _____

b) *À quelle heure te lèves-tu d'habitude?* _____

c) *L'heure de ton lever varie-t-elle d'une journée à l'autre?* _____

d) *Quelles sont les chances que d'habitude tu sois levé(e) à huit heures?* _____

Q 2. a) *À quelle heure t'es-tu couché(e) hier soir?* _____

b) *À quelle heure te couches-tu d'habitude?* _____

c) *L'heure de ton coucher varie-t-elle d'une journée à l'autre?* _____

Q 3. *Tes heures de coucher sont-elles plus régulières que tes heures de lever?* _____

Q 4. *Y a-t-il une relation entre tes heures de lever et tes heures de coucher?* _____

Expliquer cette dernière réponse:

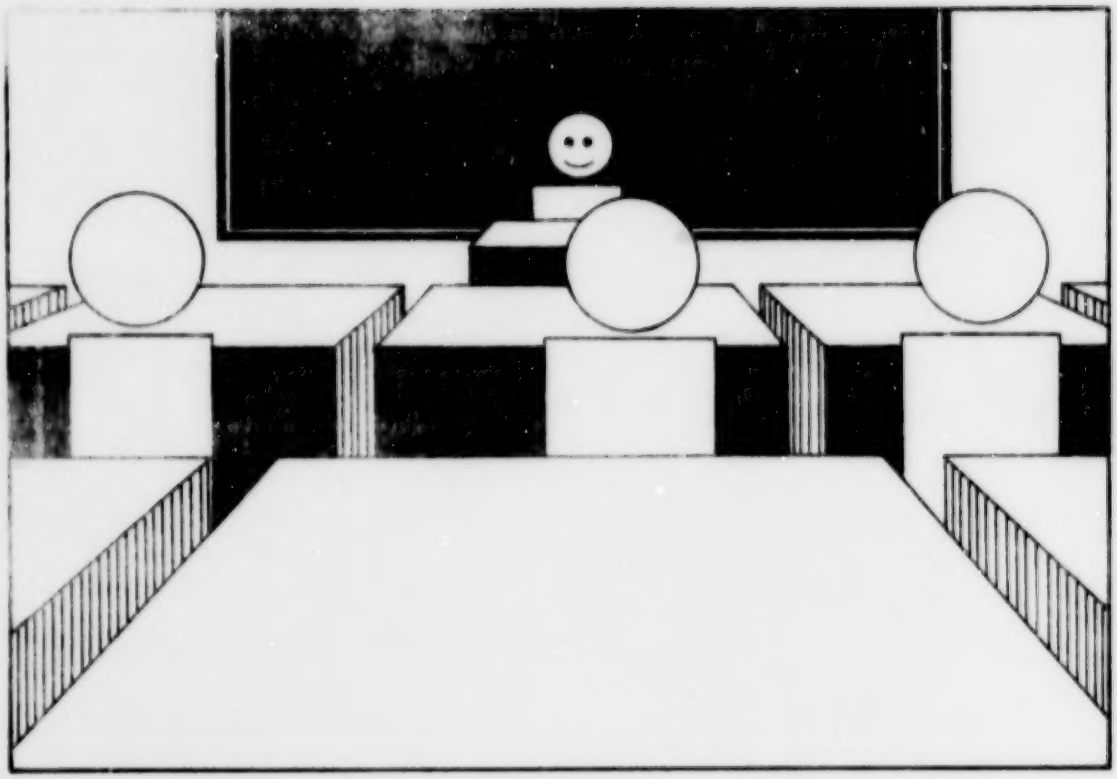
CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Coup d'oeil sur ma classe



EXPÉRIENCE N° 2

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Prendre conscience de notre connaissance intuitive des concepts des probabilités et de la statistique et de notre habilité naturelle à les véhiculer. Egalement, prendre conscience des limites de la connaissance intuitive.

MATÉRIEL

Tes compagnons de classe, tes yeux et du papier.

CONTEXTE DE L'EXPÉRIENCE

Dans l'expérience précédente, chaque étudiant(e) a décrit quelques-unes de ses habitudes; dans cette expérience, chaque étudiant(e) décrira son groupe à l'aide des outils intuitifs qu'il(elle) possède.

Partie A: Description

Décrire physiquement les étudiants(es) de la classe en une quinzaine de lignes.

Partie B: Étude particulière ou détaillée

(Répondre aux questions)

D'après toi, dans ta classe,

Q 1. Y a-t-il plus d'étudiants que d'étudiantes?

Q 3. Les étudiants sont-ils plus grands que les étudiantes?

Q 4. Les étudiants sont-ils plus âgés que les étudiantes?

Q 5. Les étudiants sont-ils plus pesants que les étudiantes?

Q 6. Quel âge les étudiant(es) ont-il(elles)?

Q 7. Quel est le poids des étudiant(es)?

Q 8. Les étudiant(es) sont-il(elles) tous(tes) de la même grandeur?

Q 9. Les étudiants sont-ils plus de même grandeur que les étudiantes?

Q 9. *Penses-tu que les étudiants(es) les plus grands(es) sont les plus pesants(es)?*

Q 10. *Ta classe est-elle une bonne représentation au point de vue physique (poids, taille)*

i) *du type d'étudiants(es) de ton institution?*

ii) *du type québécois? Expliques.*

Q 11. *Les étudiants(es) de ta classe sont-ils(elles) plus beaux(belles) ou plus laids(es) que les étudiants(es) de ton institution?*

Q 12. *Tu travailles à la Coop étudiante de ton institution et tu projettes de vendre des jeans aux étudiants(es). Qu'as-tu besoin de connaître pour faire tes achats chez le grossiste?*

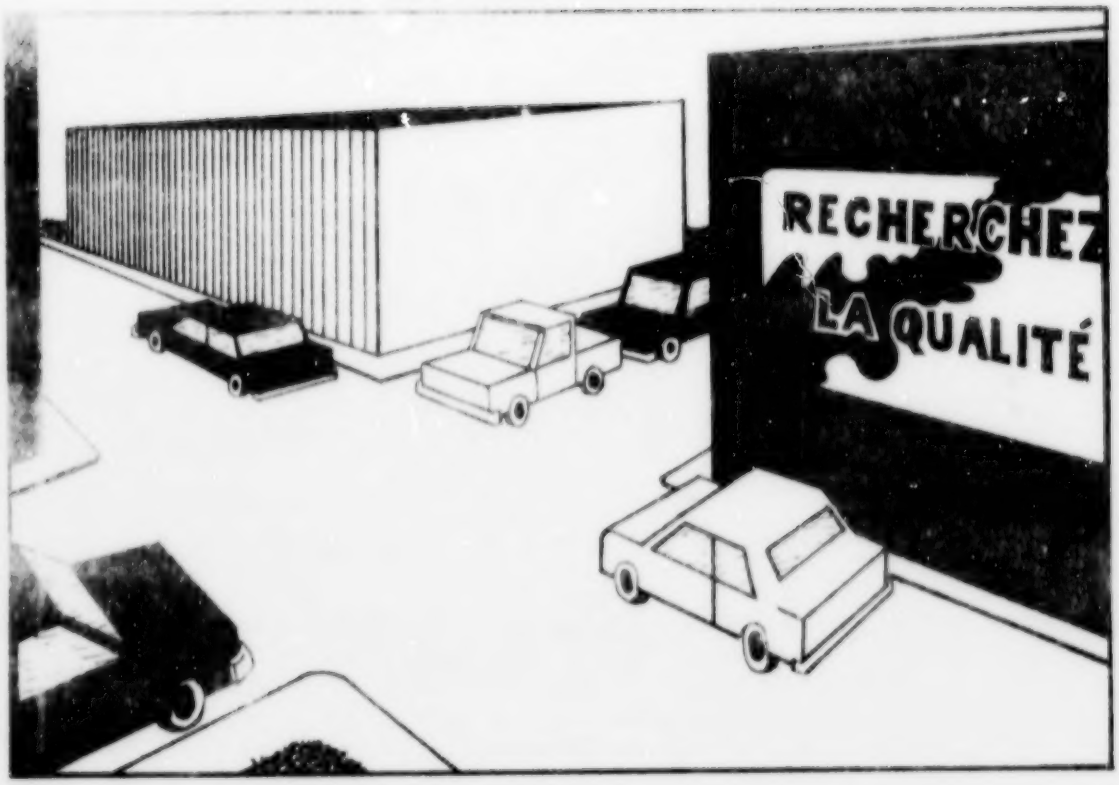
CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Observation statistique



EXPÉRIENCE P 3

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Prendre conscience de notre connaissance intuitive des concepts des probabilités et de la statistique et de notre habilité naturelle à les véhiculer. Également, prendre conscience des limites de la connaissance intuitive.

MATÉRIEL

Un crayon et du papier.

CONTEXTE DE L'EXPÉRIENCE

Dans cette expérience chaque étudiant(e) devra analyser la circulation automobile selon sa méthode propre. On pourra par la suite comparer tous ces résultats.

CHEMINEMENT DE L'EXPÉRIENCE

Tu te rends à différents endroits de la ville.

En me précisant les endroits et la durée de tes observations, tu me rapportes les informations suivantes:

ENDROIT	DUREE

Q 1. Quel est le pourcentage de conducteurs féminins?

Q 2. En général, combien y a-t-il de personnes par voiture?

Q 3. *Quelle est la densité de la circulation (nombre de voitures/heure) à ces endroits?*

Q 4. *Le pourcentage des camions est-il supérieur à 25%?*

Comment as-tu procédé pour répondre à ces questions?

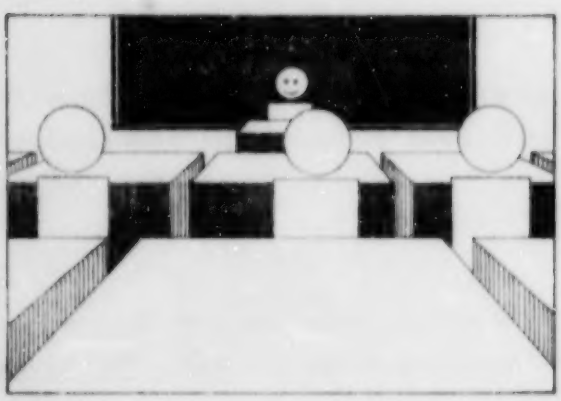
CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

DU CONNU VERS L'INCONNU
PAR LES EXPÉRIENCES PRÉALABLES



Nous allons revenir sur les expériences préalables afin d'en dégager certaines constatations et d'énumérer les thèmes des probabilités et de la statistique touchés dans ces expériences. Enfin nous reviendrons sur le besoin de raffiner les outils intuitifs pour mieux comprendre des situations rencontrées dans la vie ordinaire.

1- Constatations

À l'expérience préalable 1, chaque étudiant(e) a décrit certaines de ses habitudes de vie.

La compilation des réponses permet d'obtenir une description de ces habitudes pour l'ensemble du groupe.

À l'expérience préalable 2, chaque étudiant(e) a décrit physiquement le groupe d'étudiants(es) en utilisant sa propre méthode.

La compilation des réponses permet de comparer les différentes méthodes de description d'un groupe.

À l'expérience préalable 3, chaque étudiant(e) doit donner une idée assez juste de l'ensemble de la circulation automobile à l'aide des résultats obtenus d'une partie de cet ensemble.

La compilation des réponses donne une information plus complète sur la circulation automobile et permet de comparer les différents résultats obtenus.

2. Thèmes touchés

Dans les trois expériences préalables, nous avons touché l'ensemble des thèmes de la statistique soit:

- la méthode statistique
- la structure
- la tendance centrale
- la dispersion
- la corrélation
- les probabilités
- la théorie d'échantillonnage

Vous avez véhiculé les thèmes suivants, dans l'expérience 1:

- la tendance centrale
- la dispersion
- la corrélation
- les probabilités

dans l'expérience 2:

- la méthode statistique
- la structure
- la tendance centrale
- la dispersion
- les probabilités

dans l'expérience 3:

- la méthode statistique
- la tendance centrale
- les probabilités
- la théorie d'échantillonnage

3. Notre environnement statistique

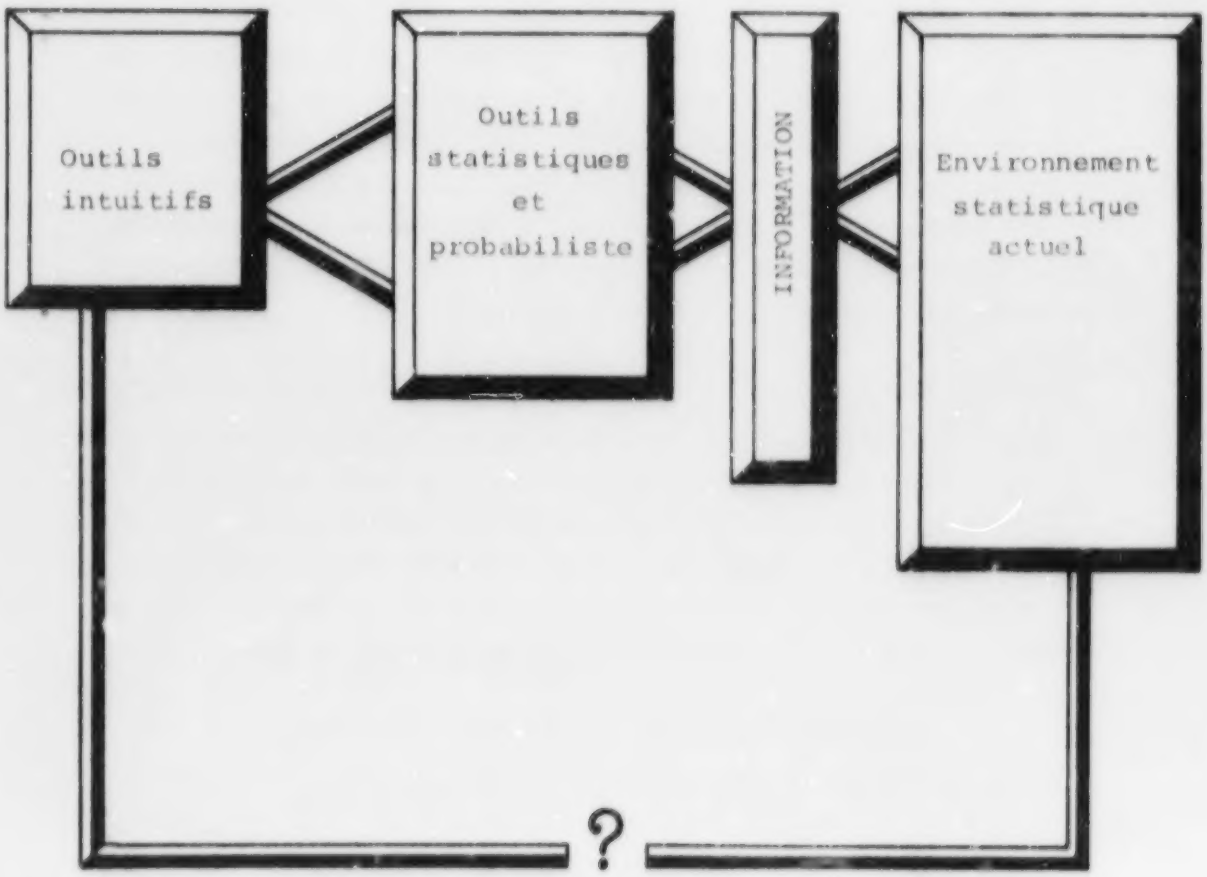
L'homme a travaillé, particulièrement depuis une cinquantaine d'années, à l'extension de ses outils intuitifs; l'essor considérable de la recherche en probabilités et en statistique en témoigne.

Les outils intellectuels, générés du fignolage des outils intuitifs, laissent leur empreinte sur tout notre environnement, dans toutes les disciplines, en agriculture, en biologie, en médecine, en psychologie, en sociologie, en génie, en administration, en linguistique, etc. et dans tous les domaines que ce soit sous forme de sondage ou d'enquête, de contrôle de la qualité, de l'expérimentation, de feuillets ou questionnaires d'apparence anodine qu'on doit répondre lors d'une demande d'installation d'un téléphone, d'une carte de crédit, etc.

Les prochaines pages représentent un petit échantillon du genre d'environnement statistique qui nous entoure. Pour comprendre cet environnement et ne pas être immergé, il devient de plus en plus utile de se sensibiliser à la méthode statistique, en comprendre les rouages, ses faiblesses, ses grandeurs et faire mentir en quelque sorte le vieil adage selon lequel: "On peut faire dire n'importe quoi à des chiffres".

Il nous faut pour cela passer de nos outils intuitifs aux outils plus rafinés qu'ils ont générés et dont on fait de plus en plus usage dans notre environnement.

La suite du cours nous amène à maftriser ces nouveaux outils dont la réussite n'est pas totalement étrangère au développement spectaculaire de l'informatique.



CARTE DE SONDAGE POINÇONNÉE

ENQUETE NO
X1 X2 X3
X4 X5 X6

RÉF NO
100
101
102

1. Vous trouverez ci-dessous une liste de raisons qui, selon les résultats de sondages précédents motivent les abonnés à lire L'Actualité. Veuillez pointer vos trois sujets préférés dans L'Actualité

( ) Sports	( ) Gens	( ) Affaires
( ) Économie	( ) Voyage	( ) Agriculture
( ) Loisirs	( ) Politiques	( ) Monde féminin
( ) Médecine	( ) Affaires étrangères	( ) Mondanités
( ) Science		

ABONNÉ(E) DEPUIS
64 69 74
65 70 75
66 71 75

2. Avant de vous abonner à L'Actualité, étiez-vous abonné à

( ) Le Maclean	( ) Aux deux
( ) Actualité	( ) Aucun

3. En lisant L'Actualité, diriez-vous que ce magazine
( ) a tendance à former l'opinion publique
( ) a tendance à refléter l'opinion publique
( ) ne l'un ni l'autre

76	77	78
79	80	81
82	83	84

A
BL
M
NP
N
NI
NS
O
PI
Q
S
YT

FÉDÉRATION QUÉBÉCOISE DE LA FAUNE

Inscription au cours de sécurité dans le maniement des armes à feu

Nom 1
Prénom
Date de naissance
Adresse 2
Sexe
Forces armées
Langue écrite
12 mois résidant-qué
Réservé au M.L.C.P.

Enquête de Profil de consommation donnant droit à des primes

C'est facile, intéressant et tout à fait confidentiel. Vous répondez au questionnaire en 10 minutes seulement.

En quoi consiste l'enquête de Profil de consommation donnant droit à des primes?

Profil de consommation est le nom d'une compagnie qui conduit des enquêtes sur les achats du consommateur au nom de plusieurs grands manufacturiers du Canada. Les résultats de cette enquête servent à établir la liste des divers produits utilisés et préférés par les consommateurs canadiens.

Que recevez-vous en retour?

Dans le cadre de cet important programme et en guise de remerciement pour avoir répondu à ce questionnaire, vous recevrez, dans les quelques jours qui suivront, un

Remplir et poster avec votre chèque ou mandat à:
LOTOMATIQUE, LOTO-QUÉBEC
C.P. 9999, Succursale Place d'Armes,
Montréal (Québec) H2Y 1A1

LOTOMATIQUE

L'ABONNE-CHANCE

CHOISISSEZ
UN OU PLUSIEURS DE CES ABONNEMENTS

Indiquer le nombre d'abonnements désiré pour chaque tirage. Chaque abonnement doit avoir un numéro différent.

TOTAL

MINI INTER PROVINCIAL SUPERLOTO 6/36	52 tirages 26 tirages 12 tirages 9 tirages 52 tirages	ABONNEMENT D'UN AN	61 ☐	*260$	× = $
MINI INTER PROVINCIAL SUPERLOTO 6/36	26 tirages 13 tirages 6 tirages 5 tirages 26 tirages	ABONNEMENT DE 6 MOIS	62 ☐	*135$	× = $
MINI		52 TIRAGES		26$	× = $
INTER		26 TIRAGES		52$	× = $
PROVINCIAL		12 TIRAGES		60$	× = $
SUPERLOTO		9 TIRAGES		90$	× = $
6/36	67 ☐ 68 ☐	52 TIRAGES 26 TIRAGES		52$ 26$	× = $ × = $

NUMÉRO DE RÉFÉRENCE **B** 5081834

6/36

Inscrire un nombre différent entre 1 et 36 dans chacune des six cases d'un même rang horizontal. Si les cases ne sont pas remplies par l'abonné, l'ordinateur de Loto-Québec émettra lui-même au sort les numéros manquants lors de l'enregistrement de l'abonnement.

N.B. N'utiliser qu'un rang horizontal de six cases par carte-jeton.

☐ FRANÇAIS ☐ ANGLAIS

GRAND TOTAL $

Faire un chèque ou mandat du montant ci-contre à l'ordre de LOTO-QUÉBEC et l'inclure avec le formulaire.

GROUPE ☐
La personne responsable du groupe remplit le présent formulaire. Elle inscrit aussi dans la section "Groupe" au verso le nom, l'adresse, le numéro de téléphone et la part de tous les abonnés.

Oui, veuillez prendre note des abonnements que j'ai choisis ci-dessus. Faites-moi parvenir au plus tôt mon certificat d'abonnement ainsi que les dates d'entrée en vigueur et d'expiration de chaque abonnement.

En remplissant et en signant ce formulaire, j'accepte les conditions inscrites sur la partie détachable que je peux conserver.

Loto-Québec se réserve le droit d'annuler ou de modifier en tout temps la totalité ou toute partie du système Lotomatique.

Nom (lettres moulées) _____ Signature _____

Adresse _____

Ville _____ Province _____ Code postal _____

Tél. résidence _____

Québec, Le Soleil, samedi 19 février 1977

La règle de Troie

Des méthodes de classement scolaire remises en cause

par Roger BELLEFEUILLE

Les méthodes de classement [illegible] des élèves des [illegible] et secondaire [illegible] un effet [illegible] sous le couvert d'une approche scientifique et souvent pour des fins **purement** administratives de confiner des centaines d'enfants au rang des "faibles" ou des "pas bons" sinon à celui de cancres institutionnalisés, indépendamment de l'effort intellectuel.

C'est du moins ce que tendent à démontrer cinq professeurs de l'Université de Montréal, dont quatre psychologues, dans une étude non encore [illegible] plus la pertinence des instruments de mesure utilisés dans les écoles comme au ministère de l'Education, rend perplexe. En particulier, les enfants issus des milieux défavorisés moins équipés pour faire face à la concurrence scolaire, risquent plus que d'autres de se retrouver éternellement dans les couches faibles de la population scolaire et, éventuellement, parmi les chroniques tâcherons de la société.

Un outil de sélection

Le modèle statistique de la courbe normale ou de la courbe de Gauss, du nom de son concepteur (Voir tableau 1) prend pour acquis au départ qu'un [illegible] handicapés physiques ou sensoriels peuvent ils être influencés par les [illegible] du hasard.

Les résultats fondés uniquement sur une approche normative et non pas diagnostique ne tiendraient pas compte dans le modèle statistique démontre, de ces facteurs qui ont souvent une influence déterminante sur le rendement scolaire d'un enfant.

Par exemple, un enfant qui subit un retard pédagogique pour diverses causes comme la maladie ou un accident pourrait fort bien, selon ce type de classement, être freiné ou limité dans son cheminement scolaire [illegible]

De plus, tous les auteurs de cette étude l'analyse de la maquette horaire de la CECM met en relief le fait qu'à de rares exceptions, un étudiant ne peut que se maintenir horizontalement ou latéralement, le cheminement ascendant lui étant à toutes fins utiles bloqué.

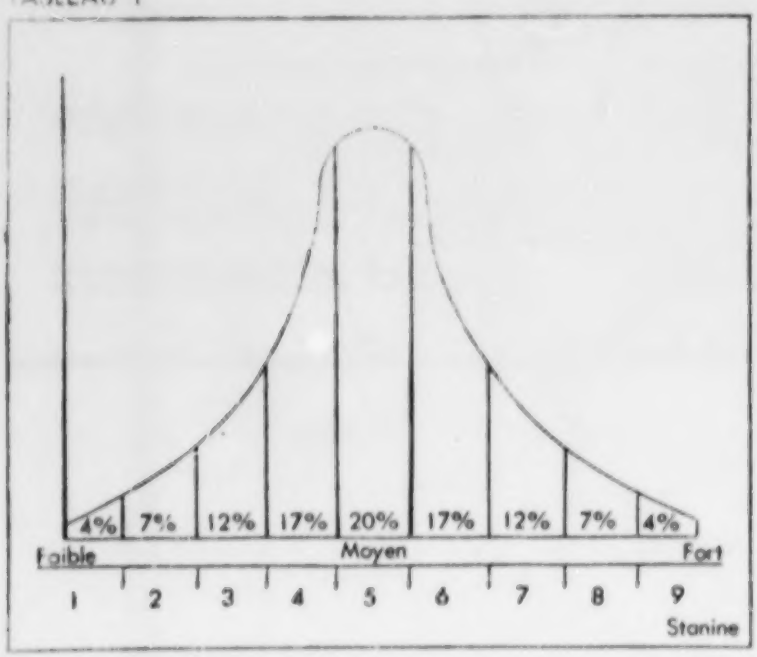
Le rang stanine de la sixième année érige donc une barrière car, la force des cours dans les concentrations principales (français et mathématiques) déterminera, soutiennent ces spécialistes de l'Université de Montréal, si un étudiant ira en professionnel court ou long, et conséquemment, s'il ira directement sur le marché du travail après le secondaire ou s'il entrera au CEGEP ou à l'université".

Fausse représentation

Les cinq auteurs soutiennent donc que les mécanismes de sélection, comme ceux appliqués à la CECM, ont des conséquences déterminantes pour l'avenir des élèves.

"Le système des tests "scientifiques" et la normalisation sont des politiques de plus en plus répandues qui étiquettent les étudiants sous de la fausse représentation", signalent-ils.

TABLEAU I



ACTIVITÉS

Les statistiques ça se comprend

POPULATION DE 15 ANS ET PLUS, POPULATION ACTIVE, TAUX D'ACTIVITE, EMPLOI ET TAUX DE CHOMAGE PAR GROUPE D'AGE ET PAR SEXE QUÉBEC, 1975

	Population de 15 ans et plus		Taux d'activité	Population active		Emploi		Taux de chômage
	'000	Importance relative %	%	'000	Importance relative %	'000	Importance relative %	%
15 à 19 ans	656	14,4	44,7	294	11,0	240	9,8	18,2
Hommes	333	14,9	48,0	160	9,2	130	8,1	18,7
Femmes	323	13,8	41,4	134	14,4	110	13,0	17,5
20 à 24 ans	601	13,2	75,8	455	17,1	404	16,5	11,3
Hommes	301	13,5	84,4	254	14,6	222	13,8	12,7
Femmes	300	12,9	67,2	202	21,6	182	21,5	9,5
25 à 44 ans	1 703	37,3	70,4	1 199	44,9	1 128	46,0	5,9
Hommes	848	38,0	95,1	807	46,5	766	47,8	5,1
Femmes	855	36,6	45,9	392	42,0	362	42,7	7,7
45 à 64 ans	1 W51	25,2	58,5	673	25,2	636	25,9	5,5
Hommes	556	24,9	86,2	480	27,7	455	28,4	5,2
Femmes	594	25,5	32,5	193	20,7	181	21,3	6,2
65 ans et plus	454	9,9	10,6	48	1,8	45	1,8	7,5
Hommes	194	8,7	18,2	35	2,0	32	2,0	—
Femmes	261	11,2	4,9	13	1,4	12	1,4	—
Total	4 565	100,0	58,5	2 669	100,0	2 452	100,0	8,1
Hommes	2 232	100,0	77,7	1 735	100,0	1 604	100,0	7,5
Femmes	2 333	100,0	40,0	933	100,0	848	100,0	9,1

Source: La situation économique au Québec - 1975.

Les tableaux statistiques nous apparaissent, pour la plupart, rebutants parce qu'ils demandent un certain effort de compréhension. Nous vous proposons donc de regarder avec nous le tableau suivant et de constater qu'il y a moyen d'y voir clair.

Comment lire le tableau

Dans la première colonne de gauche, les groupes d'âge correspondent au total des hommes et des femmes faisant partie de ce groupe. Le détail est donné en dessous lorsqu'on donne les chiffres pour les hommes puis pour les femmes.

I. POPULATION DE 15 ANS ET PLUS

A) Le symbole '000 signifie qu'il s'agit de milliers de personnes. Ainsi, il faut lire 656,000 personnes ont entre 15 ans et 19 ans.

B) Dans la colonne de l'importance relative, 14,4 signifie que les personnes âgées entre 15 ans et 19 ans représentent 14,4% de toutes les personnes âgées de 15 ans et plus.

14,9 signifie que les hommes âgés entre 15 ans et 19 ans représentent 14,9% de tous les hommes âgés de 15 ans et plus.

Commission de police du Québec

Etat de la criminalité

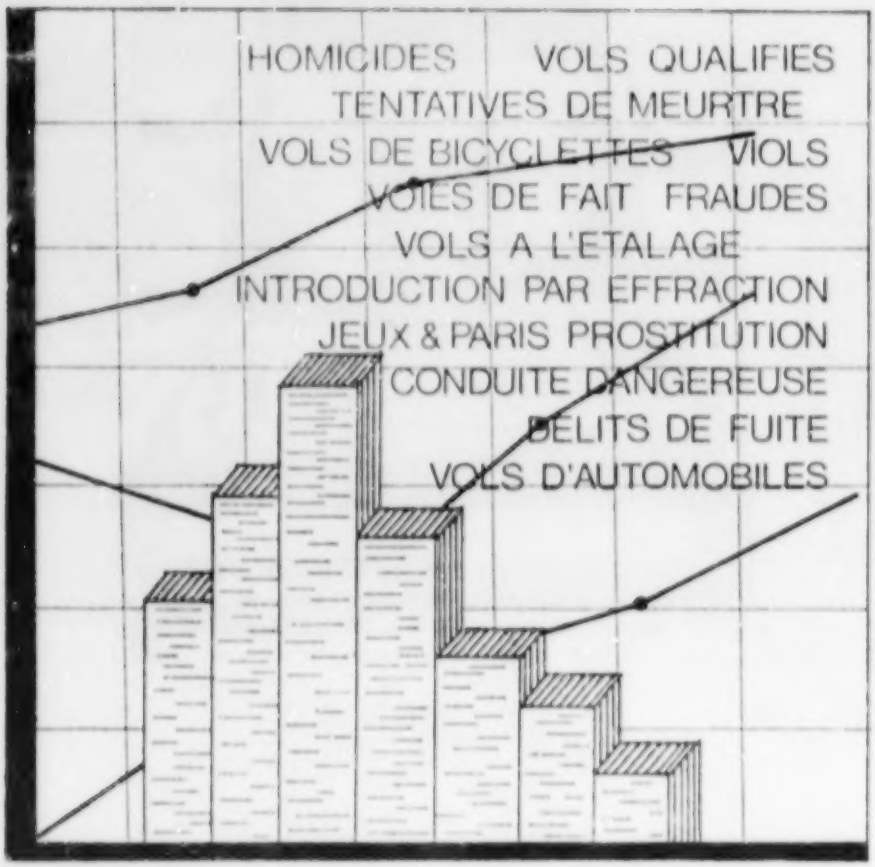
et de l'application des règlements de la circulation, infractions aux lois et autres règlements

Québec

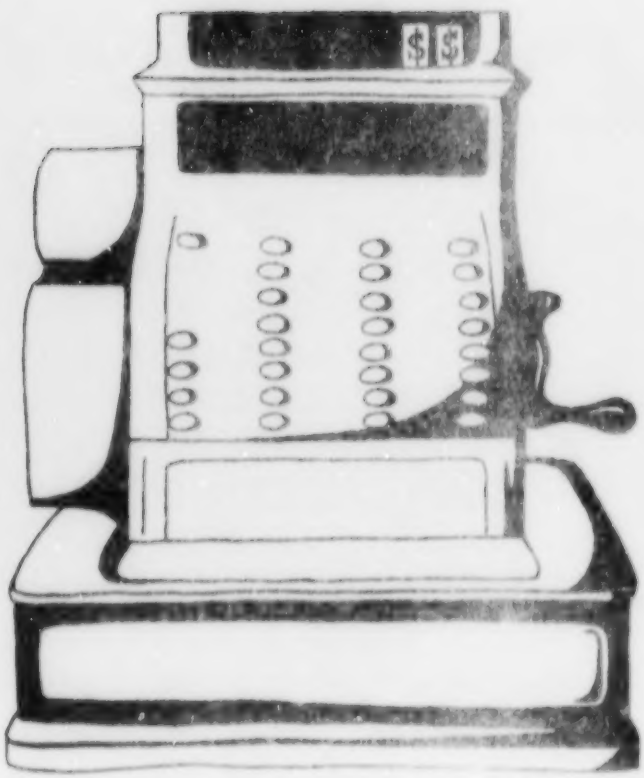
1977

Deuxième cahier

Tableaux et graphiques illustrant la statistique policière

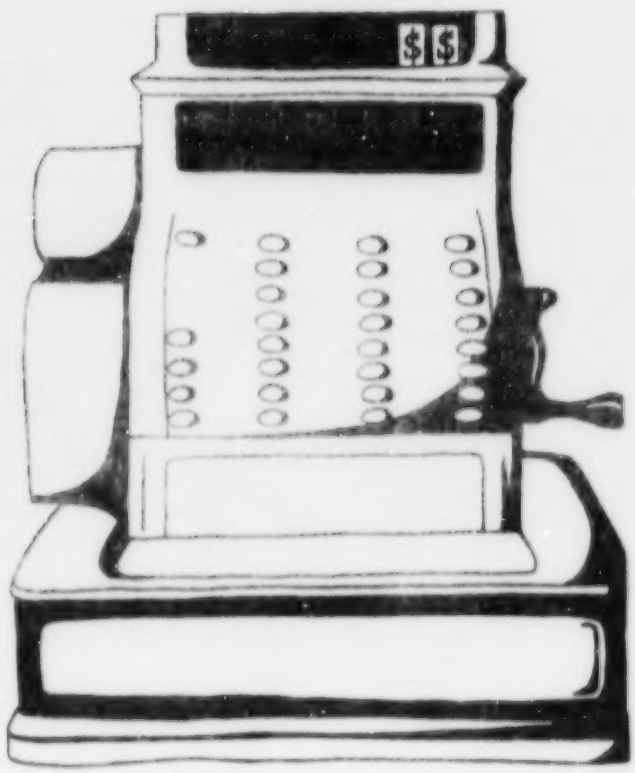


Comment le fabricant peut-il tirer profit de la statistique?



Statistique Canada Statistics Canada

Comment le détaillant peut-il tirer profit de la statistique?



Statistique Canada
Statistics Canada

Division de l'Assistance utilisateurs

Catalogue de Statistique Canada 1976-77

Revu et corrigé au 30 juin 1976

An English catalogue is available on request

Catalogue 11-204F

Partie I — Publications
Partie II — Renseignements non publiés

Publié par Statistique Canada
Publication autorisée par le ministre
de l'Industrie et du Commerce

© 2016 (sp)

ISBN 0-662-2048

Le Quotidien de Statistique Canada

TABLE DES MATIÈRES

Page

Logements mis en chantier, par genre de logements, par province, premier trimestre 1978 2

Logements mis en chantier et parachevés, mars 1978 2

Production, expéditions et stocks en mains des scieries en Colombie-Britannique, février 1978 3

Ventes d'instruments et de matériel agricoles, mars 1978 3

Ventes de télécouleurs, mars 1978 3

Situation du sucre, avril 1978 3

PARUTIONS

Éducation permanente, universités, 1976-77 (81-204, $1.00)

Opérations avec des non-résidents sur des valeurs mobilières, janvier 1978 (67-002, 40¢, $4)

Ventes de véhicules automobiles neufs, 1977 (63-208, $1.05)

Services de gaz, janvier 1978 (55-002, 30¢, $3)

Transport ferroviaire, partie III, statistique de matériel, des voies et du carburant, 1976 (52-209, 70¢)

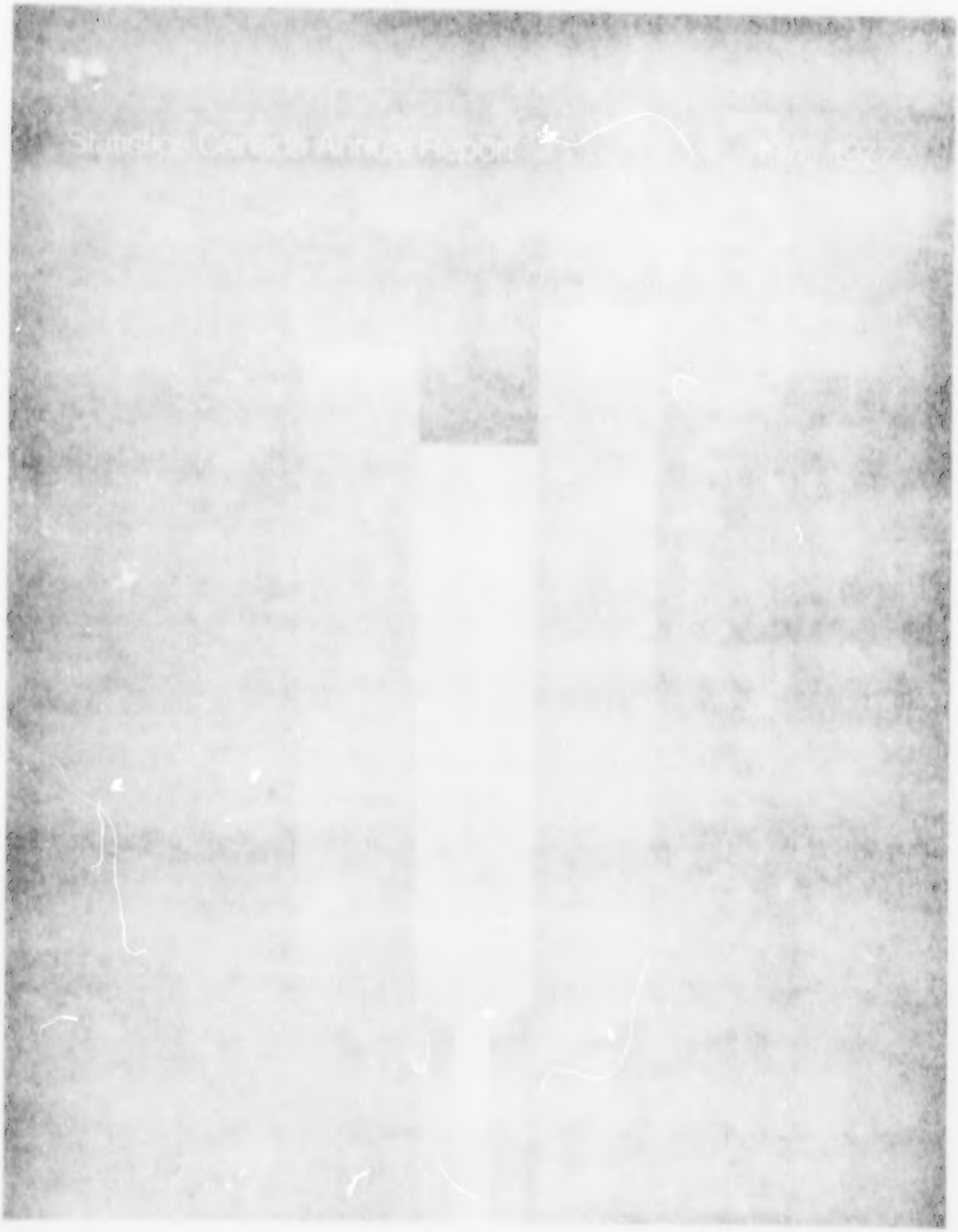
Transport ferroviaire, partie VI, statistiques de l'emploi, 1976 (52-212, 70¢)

Industrie des tapis, des carpettes et de la moquette, 1976 (34-221, 70¢)

Filature et tissage de la laine, 1976 (34-209, 70¢)

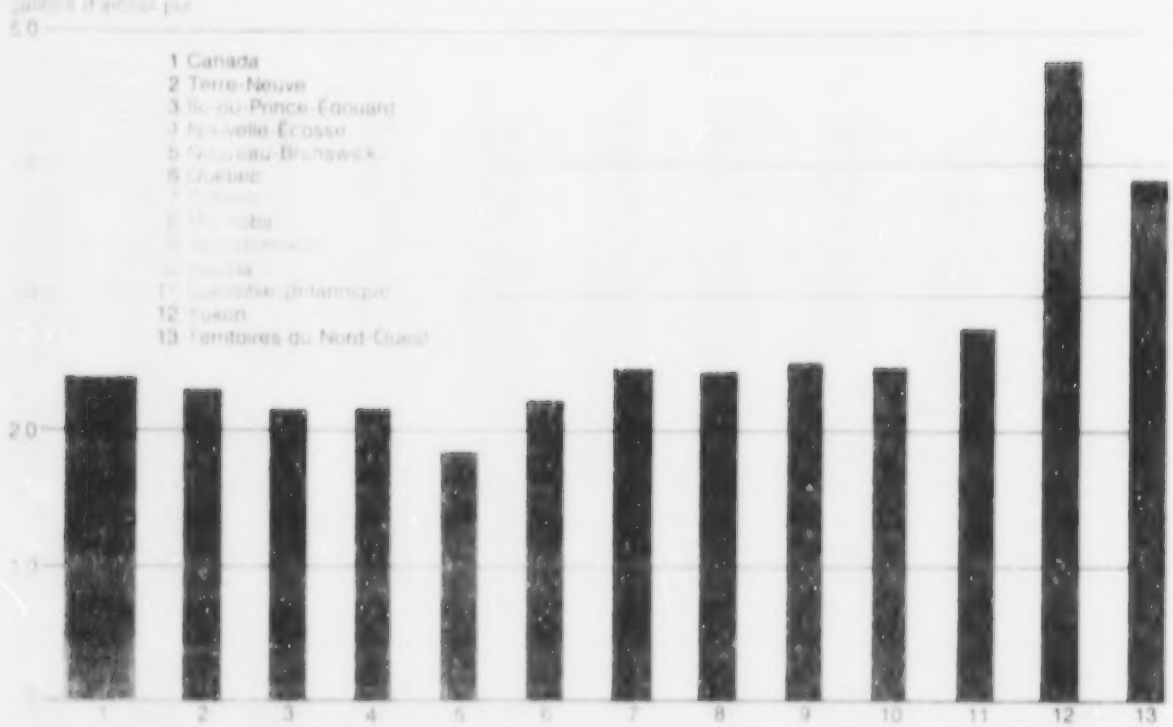
Filature et tissage du coton, 1976 (34-205, 70¢)

Stocks de viandes congelées, avril 1978 (32-012, 30¢, $3)



Graphique 4.9

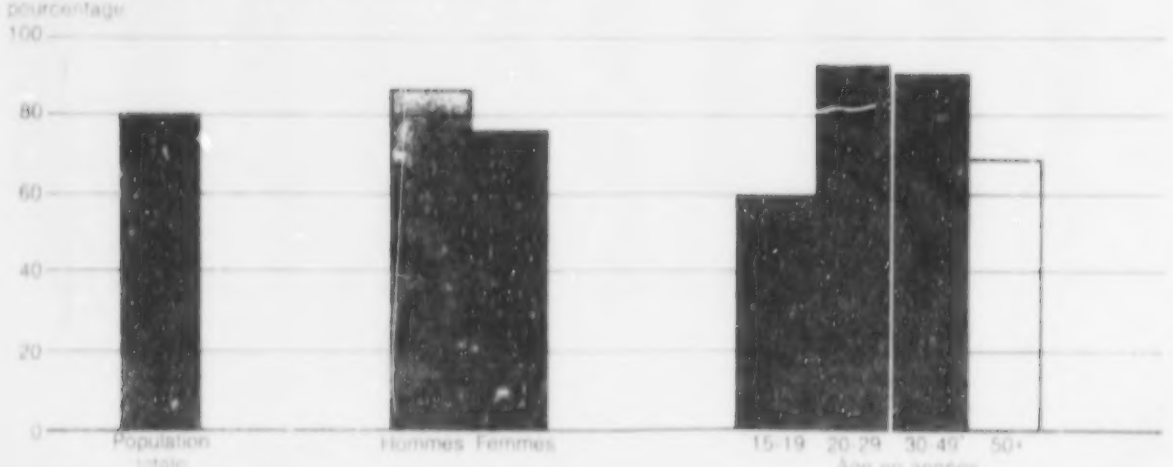
CONSOMMATION D'ALCOOL PUR PAR PERSONNE ÂGÉE DE 15 ANS ET PLUS SELON LA PROVINCE, 1974(1)



(1) Représente l'année financière terminée le 31 mars 1974

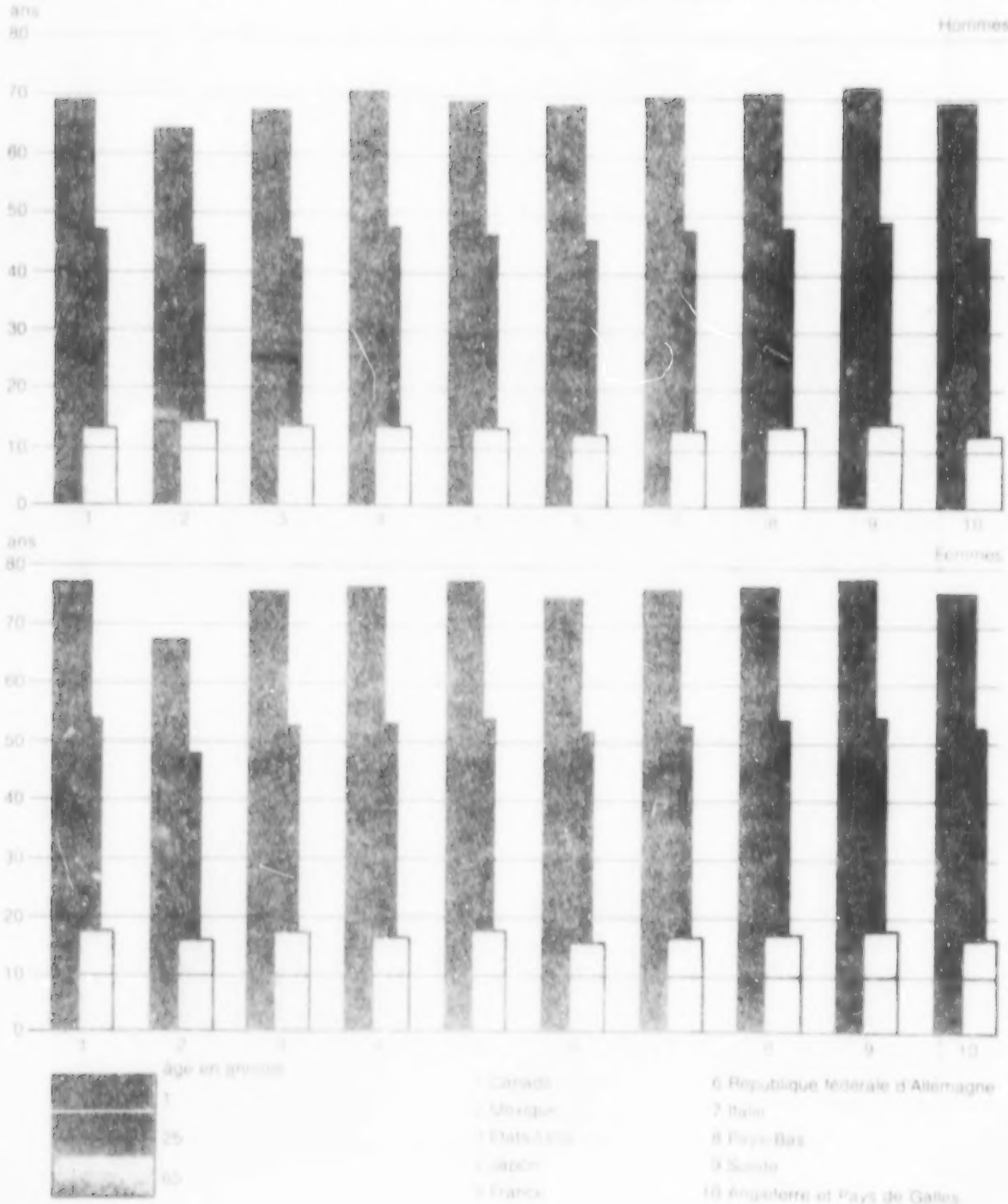
Graphique 4.10

PROPORTION DE LA POPULATION DE L'ONTARIO ÂGÉE DE 15 ET PLUS QUI CONSOMME DE L'ALCOOL, SELON L'ÂGE ET LE SEXE, 1969



Graphique 4.25

ESPÉRANCE DE VIE À CERTAINS ÂGES POUR CERTAINS PAYS, SELON LE SEXE, 1972



1 Canada
2 Norvège
3 États-Unis
4 Japon
5 France
6 République fédérale d'Allemagne
7 Italie
8 Pays-Bas
9 Suède
10 Angleterre et Pays de Galles

117

Tableau 6.9

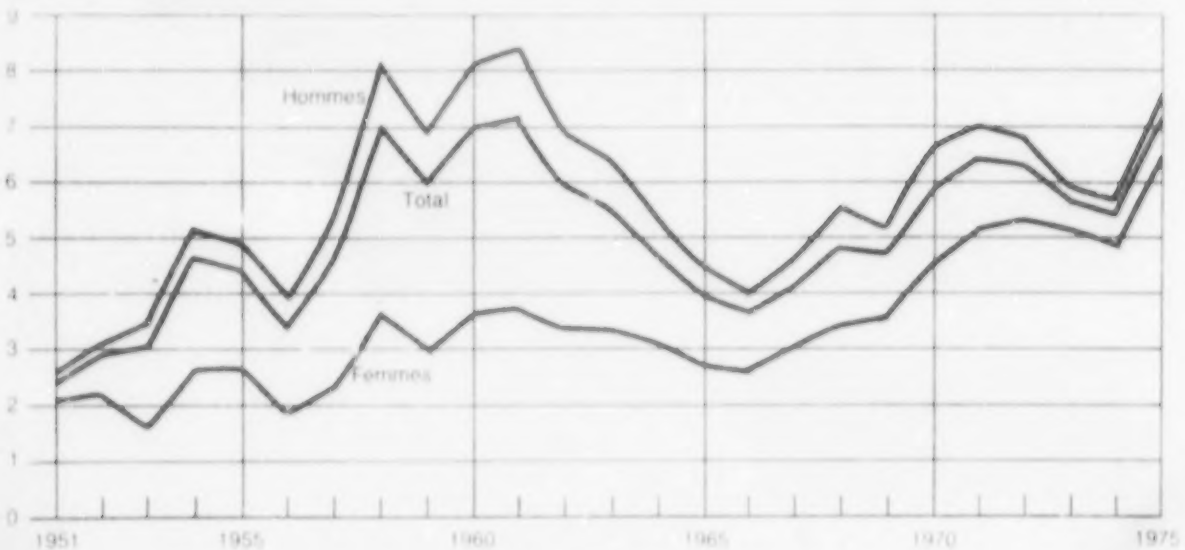
PERSONNES OCCUPÉES SELON LA RÉGION ET LE SECTEUR D'ACTIVITÉ 1975

	Atlantique		Québec		Ontario		Prairies		Colombie-Britannique	
	milliers	%	milliers	%	milliers	%	milliers	%	milliers	%
Industries primaires:										
Agriculture	22	3.1	88	3.6	121	3.4	223	14.6	26	2.5
Autres industries primaires(1)	37	5.2	46	1.8	57	1.6	46	3.0	43	4.2
Total, industries primaires	59	8.3	134	5.4	178	5.0	269	17.6	69	6.7
Industrie manufacturière	106	14.9	809	32.6	915	25.6	161	10.5	165	16.1
Construction	59	8.3	155	6.1	217	6.1	102	6.7	77	7.5
Transports, communications et autres services publics	75	10.6	211	8.6	274	7.6	143	9.4	104	10.1
Commerce	135	19.0	413	16.8	616	17.2	276	18.1	193	18.8
Finances, assurances et affaires immobilières	24	3.4	114	4.6	193	5.4	67	4.4	57	5.6
Services communautaires, commerciaux et personnels	195	27.5	671	27.3	946	26.4	401	26.2	296	28.8
Administration publique	57	8.0	160	6.5	244	6.8	111	7.3	67	6.5
Total	710	100.0	2 462	100.0	3 581	100.0	1 528	100.0	1 027	100.0

(1) Comprend l'agriculture, la sylviculture, la pêche et le piégeage ainsi que les mines, carrières et puits de pétrole.

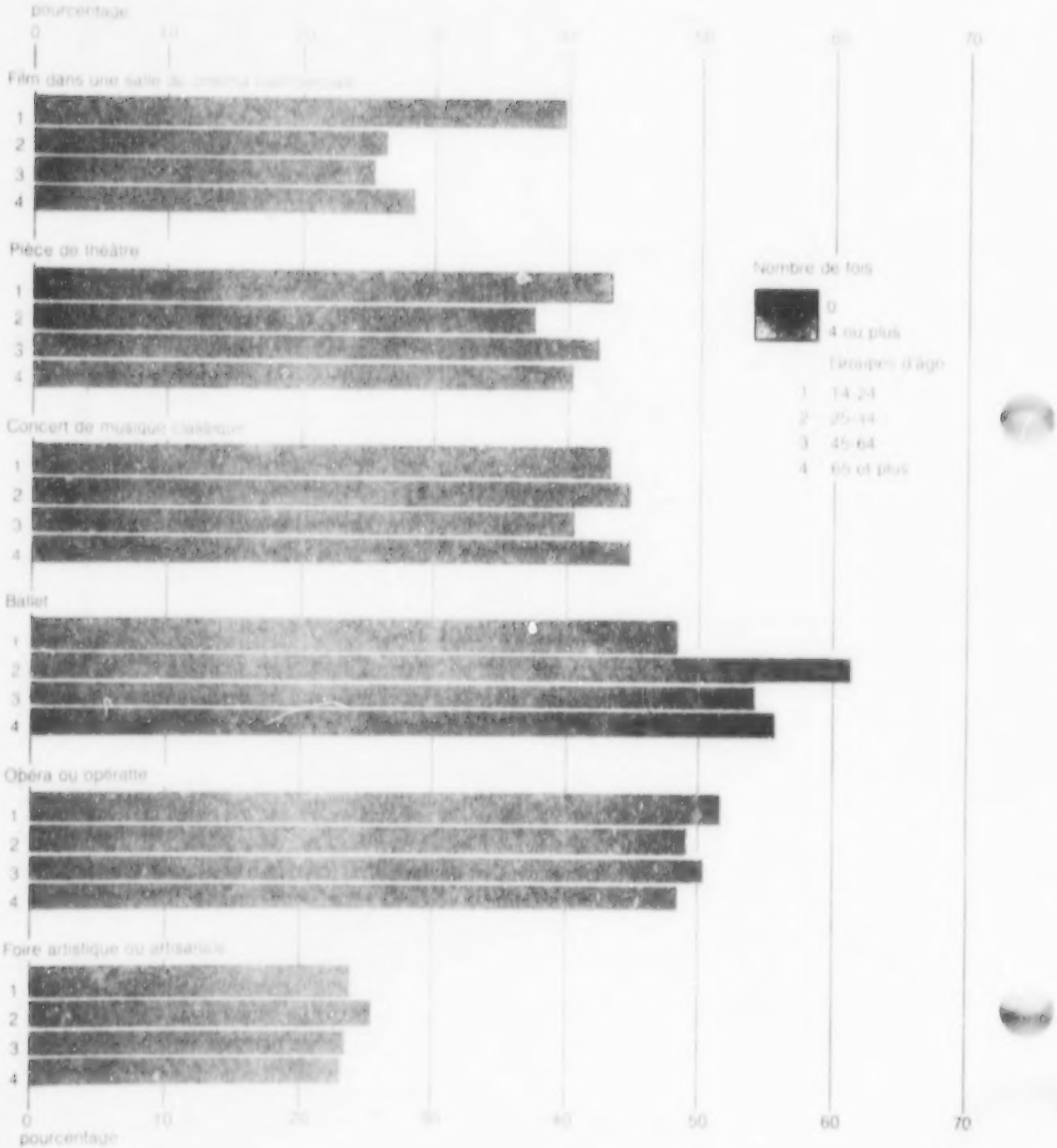
Graphique 6.10

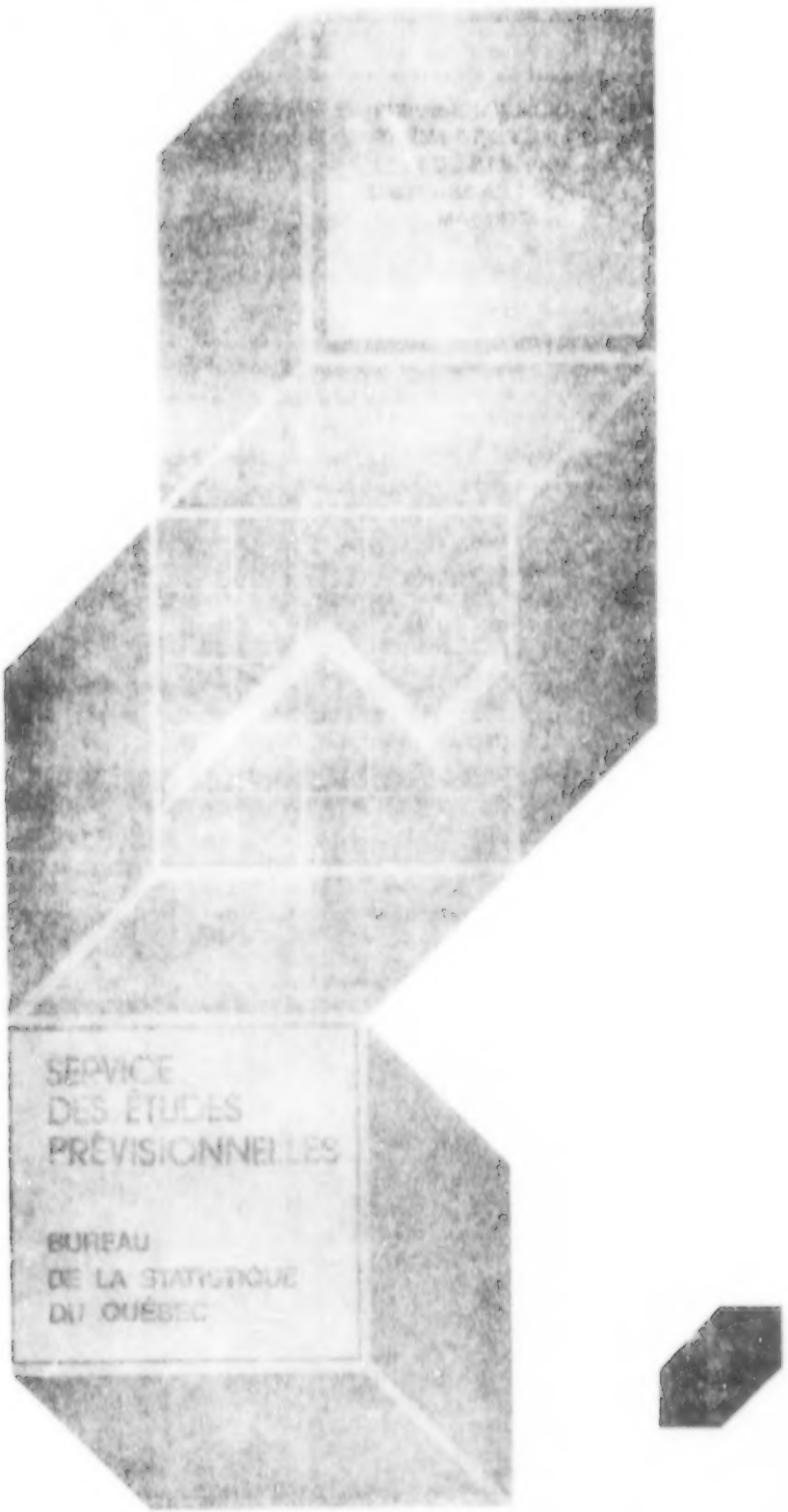
TAUX DE CHÔMAGE SELON LE SEXE



Graphique 7.14

ASSISTANCE À CERTAINS SPECTACLES RÉCRÉATIFS ET CULTURELS SELON L'ÂGE, JUILLET ET AOÛT, 1975




SERVICE
DES ÉTUDES
PRÉVISIONNELLES
BUREAU
DE LA STATISTIQUE
DU QUÉBEC

BUREAU DE LA STATISTIQUE DU QUÉBEC

Direction de la diffusion et de l'assistance technique

SERVICE DES METHODES

LOCALISATION DANS LA DIRECTION

PROBLÉMATIQUE GÉNÉRALE

Dans le cadre de [illegible] organismes publics et parapublics, on retrouve les domaines spécifiques de la statistique suivants pour lesquels [illegible] une assistance :

1. la méthodologie d'enquête ;
2. l'analyse statistique des données ;
3. la standardisation des codes ;
4. la normalisation dans les [illegible] ;

Devant la nécessité du développement [illegible] méthodes statistiques et en standardisation [illegible] organismes publics et parapublics, [illegible] domaines spécifiques de la statistique [illegible] une assistance de haut calibre dans [illegible] ; c'est la raison d'être des méthodes. Ce dernier se [illegible] deux divisions :

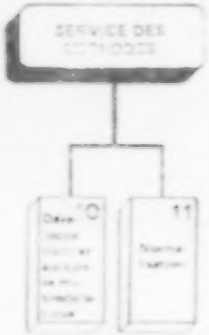
- la Division du développement et de l'assistance méthodologique [illegible] répondre aux besoins relatifs à la méthodologie d'enquête et l'analyse statistique des données ;
- la Division de la normalisation [illegible] répondre aux besoins de standardisation et de [illegible].

TABLEAU DES EFFECTIFS

Catégories	En nombre	En %
Cadres	0	–
Professionnels	6	42
Autres	8	57
Total	14	100

① D'après la répartition en novembre 1976

LA STRUCTURE ORGANISATIONNELLE



UNITÉS	MANDATS	OUTPUTS
10	● Participer à la conception et à la réalisation de méthodologies d'enquête ; ● Assister et conseiller dans l'utilisation des méthodes statistiques et mathématiques ; ● Soutenir dans l'évaluation de la fiabilité des séries statistiques ; ● Se tenir à l'affût des nouveaux développements méthodologiques et analytiques.	▸ Plans de sondage ; ▸ Estimations de variance ; ▸ Méthodes d'imputation ; ▸ Contrôles de qualité ; ▸ Analyse statistique ; ▸ Nouvelles techniques méthodologiques et analytiques.
11	● Adapter, réviser et diffuser des systèmes de codification et de classification de données statistiques ; ● Être disponible à titre de consultant et apporter un soutien auprès des utilisateurs de tout système de codification ou de classification ; ● Servir de canal privilégié d'information et de discussion auprès des autres organismes de normalisation statistique.	▸ Classifications nouvelles et/ou révisées ; ▸ Atlas du code géographique du Québec ; ▸ Fichiers nouveaux et/ou révisés.

TRAVAIL DE RECHERCHE

Les tableaux et les graphiques que nous avons examinés jusqu'ici illustrent quelques sujets traités à l'aide des probabilités et de la statistique.

Recherche en bibliothèque ou ailleurs trois sujets différents et propres à ton champ d'étude (spécialité ou concentration) qui font ou ont déjà fait l'objet d'une étude statistique.

Inclue dans ton dossier des photocopies illustrant les sujets et indique les références.

Laboratoire expérimental

LABORATOIRE EXPÉRIMENTAL
OBJECTIFS
CONCEPTS
OUTILS
RÉALISATION
1) Cognitifs
- Construire quelques outils de description de données.
- Couvrir la majorité des thèmes des probabilités et de la statistique.
2) Comportemental
- Sensibiliser l'étudiant(e) aux principaux concepts des probabilités et de la statistique.
- Caractère et catégorie
- Classification et ordonnancement
- Effectif et fréquence
- Représentation graphique
- Distribution
- Tendance centrale
- Dispersion
- Lien entre deux caractéristiques
- Population et échantillon
- Méthode d'échantillonnage
- Liens entre échantillon et population
- Expérience aléatoire
- Espace échantillon - équiprobabilité
- Probabilité d'un événement
- Événement composé
- Dépendance et indépendance
- Comptabilité et incompatibilité
- Événement conditionnel
- Erreur d'échantillonnage
- Tableau d'observations
- Tableau d'effectifs
- Représentation graphique
- Torve de distribution
- Moyenne arithmétique
- Mode - Médiane
- Étendue - Écart-moyen
- Covariance
- Droite de régression
- Échantillon
- Tendance et dispersion des moyennes d'échantillons
- Probabilité
- Distribution d'échantillonnage des moyennes
- Probabilité
Expérience 1
- Caractéristiques et catégories
Expérience 2
- Classification
Expérience 3
- Distribution
Expérience 4
- Tendance centrale
Expérience 5
- Dispersion
Expérience 6
- Corrélation
Expérience 7
- Population et échantillon
Expérience 8
- Échantillonnage
Expérience 9
- Probabilité
Expérience 10
- Erreur d'échantillonnage

Caractéristiques et catégories

EXPÉRIENCE 1

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Faire la description d'un groupe de bonshommes sous forme de tableau d'observations après avoir sélectionné des caractéristiques pertinentes à cette description.

MATÉRIEL

Un groupe de 20 bonshommes, une plaque de fixation et une règle.

CONTEXTE DE L'EXPÉRIENCE

Dans cette première expérience vous devrez décrire un groupe de bonshommes. Nous vous proposons une méthode qui consiste à rechercher un certain nombre de caractéristiques du groupe et à décrire chaque bonhomme par rapport à ces caractéristiques.

CHEMINEMENT DE L'EXPÉRIENCE

Piger 20 bonshommes d'une boîte commune.

Disposer les bonshommes sur la plaque.

Faire un tableau d'observations de la façon suivante:

a) *Énumérer au moins six autres caractéristiques des bonshommes:*

1. *Sexe*
2.
3.
4.
5.
6.
7.
8.
9.
10.

b) *Remplir le tableau d'observations:*

1. *Écrire les caractéristiques sur la première ligne.*
2. *Indiquer les catégories des bonshommes suivant les différentes caractéristiques, par exemple masculin ou féminin pour le sexe.*

TABLEAU D'OBSERVATIONS

Caractéristique / Individu	SEXE									
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										

Q 1. *Si tu te réfères au tableau d'observations, en quoi le dixième bonhomme est-il distinct de la plupart des autres?*

Q 2. *En quoi le dixième bonhomme est-il semblable à la majorité des autres?*

CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Classification

RÉPARTITION DES QUÉBÉCOIS SELON
LA LANGUE D'USAGE EN 1971

EXPÉRIENCE 2

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Commencer l'étude de la structure d'une caractéristique en cherchant sa distribution et la présenter par un tableau et un graphique.

MATÉRIEL

Un groupe de 20 bonshommes, une plaque de fixation munie d'un système d'axes et une règle.

CONTEXTE DE L'EXPÉRIENCE

Dans la première expérience, vous avez décrit le groupe en examinant chaque individu suivant plusieurs caractéristiques. Toutefois, cette forme de description, dans certains cas, n'est pas utile.

Par exemple, supposons que vous devez acheter des skis pour 20 personnes. Comme la longueur des skis est principalement déterminée par la taille des individus, il vous faudrait analyser ces personnes uniquement d'après cette caractéristique. En d'autres termes, il n'est pas important de savoir que Monsieur X a les cheveux bruns, s'habille de telle façon, etc. mais combien il y a d'individus pour chaque taille différente.

Dans cette deuxième expérience, vous devrez décrire votre groupe suivant une seule caractéristique à la fois en cherchant sa distribution.

CHEMINEMENT DE L'EXPÉRIENCE

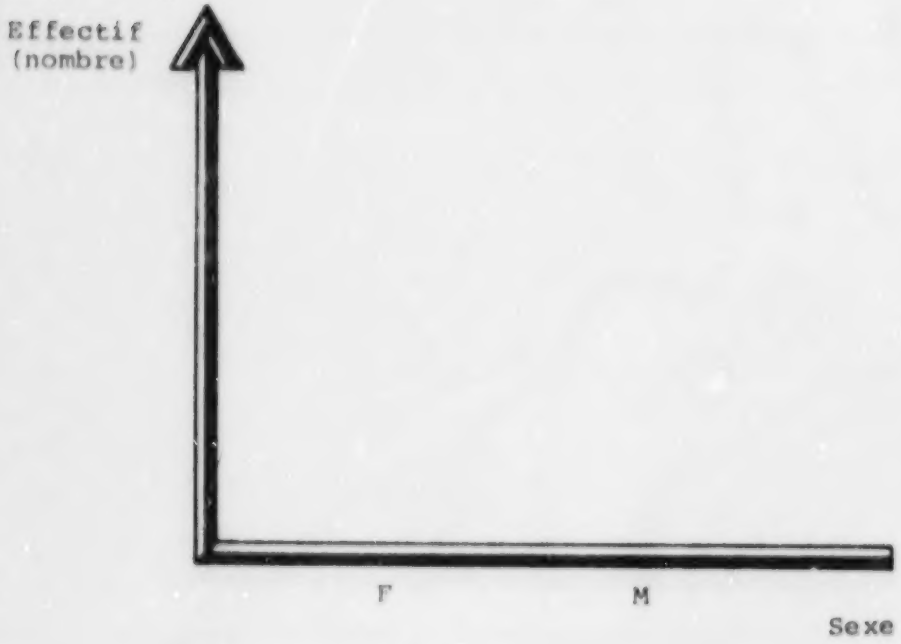
Regrouper à la file indienne sur la plaque les bonshommes suivant leur sexe (une file par catégorie et un espace constant entre chaque bonhomme).

Indiquer le nombre de cas masculins et féminins.

Sexe	Nombre
M	
F	
Total	

On a ainsi représenté la distribution de la caractéristique sexe par un tableau d'effectifs.

Faire une représentation graphique de la distribution de la caractéristique sexe.



Q 1. Votre graphique ressemble-t-il à l'image des bonshommes vus du dessus?

Mentionner s'il y a lieu les points communs entre les deux.

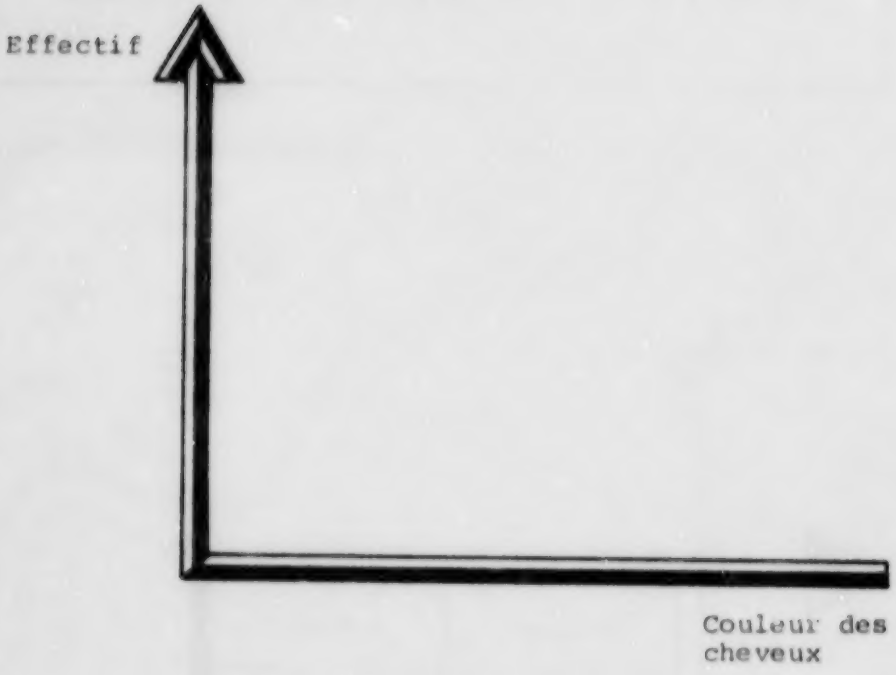
2.

Regrouper, à la file indienne, sur la plaque les bonshommes suivant leur couleur de cheveux (ne plus tenir compte du sexe).

Faire le tableau d'effectifs.

Couleur des cheveux	Effectif

Faire une représentation graphique.



Q 8. *Y a-t-il concordance entre votre montage et votre représentation graphique? Expliquer.*

3.

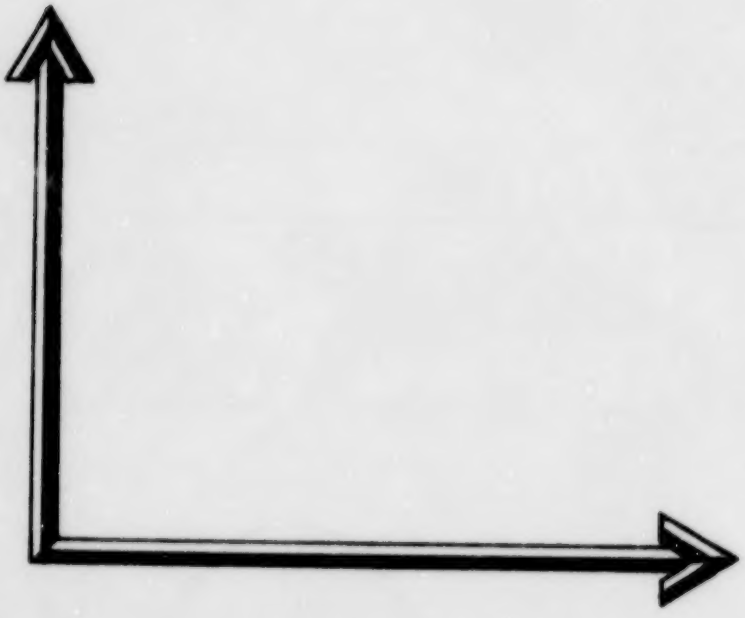
Disposer les bonshommes sur la plaque suivant leur taille classée par ordre croissant (de la plus petite à la plus grande) en regroupant les individus d'une même taille à la file indienne.

Faire le tableau d'effectifs.

Taille	Effectif

Faire une représentation graphique.

Effectif



Observer la ressemblance entre votre représentation graphique et votre montage vu du dessus.

L'effectif de chaque taille différente est représenté par une file sur le montage et une colonne sur le graphique.

La longueur d'une file est proportionnelle à la hauteur de la colonne correspondante.

CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Distribution

EXPÉRIENCE 3

Nom(s): _____

_____

OBJECTIF GÉNÉRAL

Se senbiliser à la reconnaissance de différentes formes de distribution en particulier celle de la distribution normale.

MATÉRIEL

Cinq groupes de bonshommes sur une plaque, une règle.

CONTEXTE DE L'EXPÉRIENCE

Les deux premières expériences vous ont permis d'amorcer la description de groupes de bonshommes: choix de variables, cueillette de données, classement, représentations graphiques.

On a constaté également à l'expérience 2 la nécessité de connaître la distribution d'une caractéristique pour répondre à un problème donné.

L'expérience actuelle permettra de dégager des critères qui facilitent la reconnaissance et la comparaison des distributions.

CHEMINEMENT DE L'EXPÉRIENCE

Ton montage représente des groupes classifiés suivant la taille des individus.

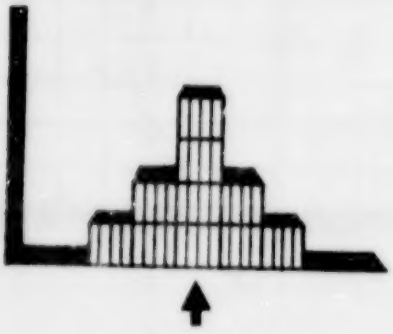
Regarder ce montage à vol d'oiseau.

Dessiner le contour de chaque groupe dans le tableau 3.1 de la page E 3.6.

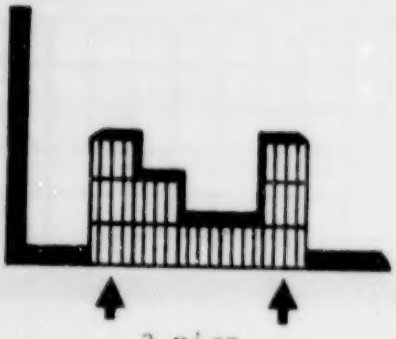
Q 1. Quels sont les groupes qui ont un contour semblable?

_____

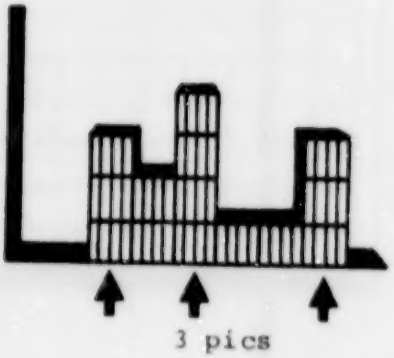
Les contours suivants présentent un certain nombre de pics.



1 pic



2 pics



3 pics

Répondre aux questions 2, 3 et 4 dans le tableau 3.2 de la page E 3.7

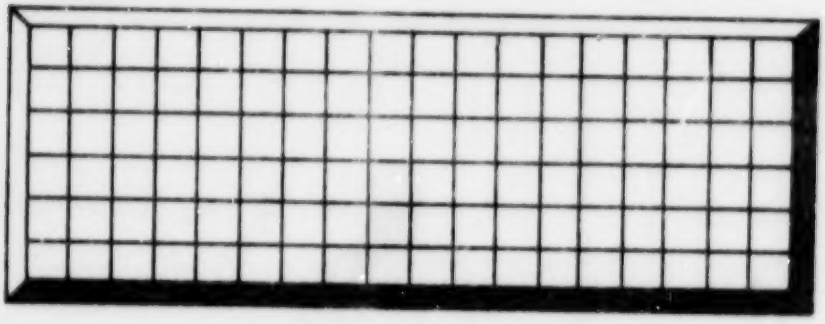
Q 2. *Combien de pics retrouve-t-on pour chacun des groupes de votre montage?*

Q 3. *Où se trouvent ces pics? (Cocher la case appropriée).*

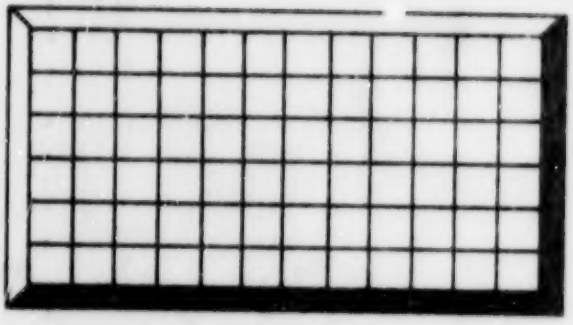
Q 4. *Quels sont les groupes symétriques par rapport au pic central? (Cocher).*

Tableau 3.1

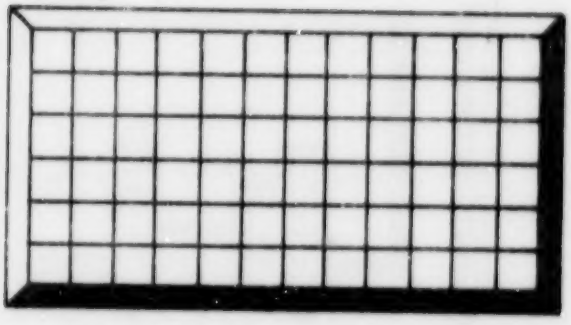
Groupe 5



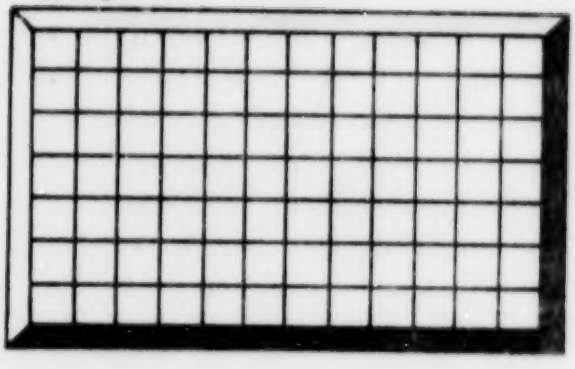
Groupe 3



Groupe 4



Groupe 1



Groupe 2

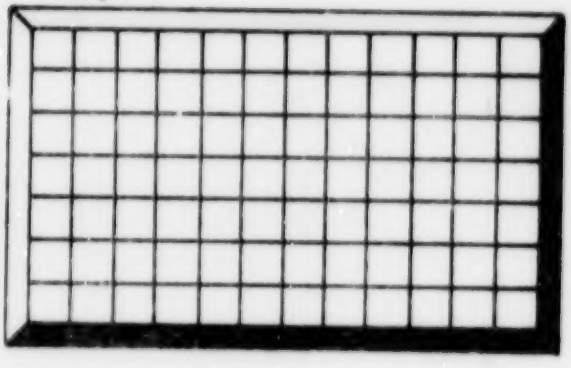


Tableau 3.2

QUESTION \ GROUPE		1	2	3	4	5
Q 2.	Nombre de pics					
Q 3.	Au centre					
	À l'extrémité gauche					
	À l'extrémité droite					
Q 4.	Symétrie par rapport au pic central					

Q 5. Quel est le pourcentage d'individus qui ont une taille:

	Groupe 4	Groupe 5
i) Égale au pic central?		
ii) Égale au pic central plus ou moins 2 unités?		
iii) Égale au pic central plus ou moins 4 unités?		
iv) Égale au pic central plus ou moins 6 unités?		

Q 6. La répartition des tailles dans le groupe 4 est-elle différente de celle du groupe 5?

_____

Expliquer _____

_____

_____

_____

_____

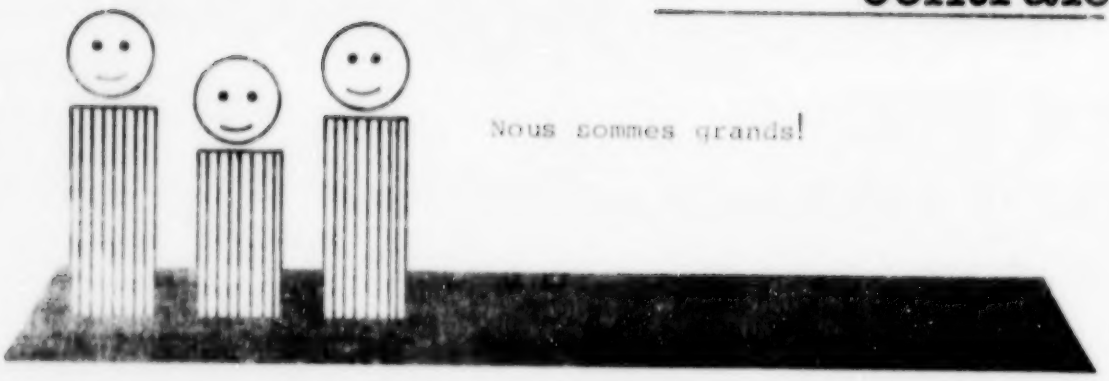
CONCLUSION

Rappeler le cheminement de l'expérience.

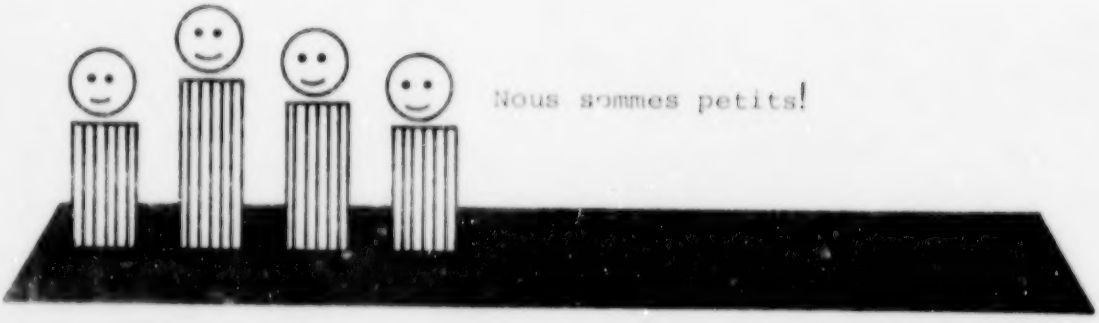
Mentionner les principales constatations.

Ton appréciation:

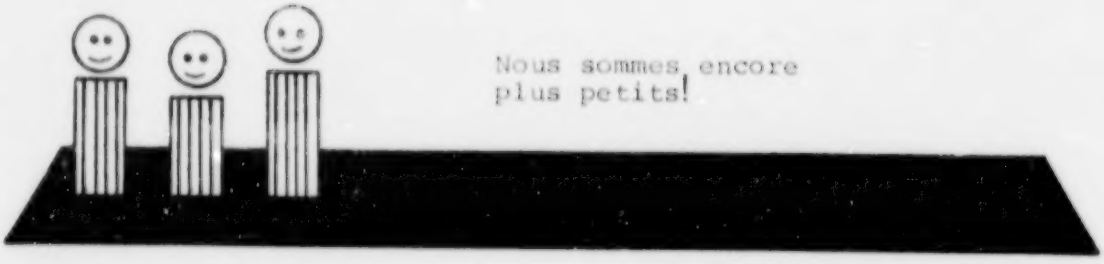
Tendance centrale



Nous sommes grands!



Nous sommes petits!



Nous sommes encore plus petits!

EXPÉRIENCE 4

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Se sensibiliser à la tendance centrale et à sa mesure.

MATÉRIEL

Neuf groupes de bonshommes disposés sur une plaque, une règle.

CONTEXTE DE L'EXPÉRIENCE

Les deux premières expériences vous ont permis d'amorcer la description de groupes de bonshommes: choix de caractéristiques, cueillette de données, classement, graphiques.

La troisième expérience vous a permis de sortir des critères servant à poursuivre la description et à comparer les groupes.

La présente expérience permet de continuer la description et la comparaison entre les groupes en utilisant une valeur (ou des valeurs) qui représente(nt) l'ensemble des données du groupe.

CHEMINEMENT DE L'EXPÉRIENCE

Partie A: Approche intuitive

Tu disposes de 9 groupes de bonshommes (les groupes sont restreints dans le but de faciliter le travail).

Questions (répondre en choisissant le symbole approprié $<$, $>$, $=$).

Q 1. Les tailles des bonshommes du groupe 1 sont en général ---3---------- les tailles des bonshommes du groupe 3.

Q 2. Les tailles des bonshommes du groupe 1 sont en général ---3---------- les tailles des bonshommes du groupe 4.

Q 3. Les tailles des bonshommes du groupe 1 sont en général ---3---------- les tailles des bonshommes du groupe 2.

Classer les groupes à partir de celui où la taille générale est la plus petite jusqu'à celui où cette taille est la plus grande. Indiquer au tableau 4-5 sous le numéro du groupe le rang qu'il occupe dans le classement.

Indiquer ici les critères employés pour faire ce classement.

Partie B: Outils proposés

Un premier outil pourrait consister à additionner la taille de tous les individus d'un groupe.

Mesurer la taille de chaque bonhomme et l'indiquer dans les cases correspondantes du tableau 4-1.

Faire la somme des tailles de chaque groupe (répondre au tableau 4-1).

TABLEAU 4-1

Groupe	Tailles	Somme des tailles
1		
2		
3		
4		
5		
6		
7		
8		
9		

En examinant les tailles des groupes 1 et 2, on constate que les bonshommes du groupe 1 sont généralement plus grands que ceux du groupe 2 mais que la somme de leurs tailles est plus petite, parce qu'ils sont moins nombreux.

Le résultat de ce premier calcul ne donne rien de significatif parce que la somme dépend de deux facteurs, soit la taille des individus et leur nombre.

Étant donné que c'est seulement la tendance de la taille qui nous intéresse, nous allons répartir la somme obtenue sur le nombre d'individus qui ont contribué à la donner.

Diviser les sommes obtenues par le nombre de bonshommes (répondre au tableau 4-2).

TABLEAU 4-2

Groupe numéro	1	2	3	4	5	6	7	8	9
somme / nombre									

La valeur de taille ainsi obtenue est en quelque sorte indépendante du nombre d'individus qu'il y a dans le groupe.

Un deuxième outil utilisable pour évaluer la taille représentative d'un groupe pourrait consister à prendre une valeur qui sépare en deux parties égales le groupe de bonshommes ordonnés selon leur taille.

Remplir l'espace approprié ci-dessous.

TABLEAU 4-3

Groupe numéro	1	2	3	4	5	6	7	8	9
Taille au centre									

Un troisième outil utilisable pour évaluer la taille représentative d'un groupe serait de choisir la(les) valeur(s) de taille qui revient(reviennent) le plus souvent dans chacun des groupes.

Remplir l'espace approprié ci-dessous.

TABLEAU 4-4

Groupe numéro	1	2	3	4	5	6	7	8	9
Taille(s) la(les) plus fréquente(s)									

Partie C: Synthèse

Reporter au tableau 4-5 les valeurs obtenues avec chacun des trois outils.

Ordonner les groupes du plus petit au plus grand selon chacun de ces outils (réponare au tableau 4-5).

Tu peux maintenant examiner la correspondance entre l'ordre intuitif et l'ordre obtenu avec chacun des outils.

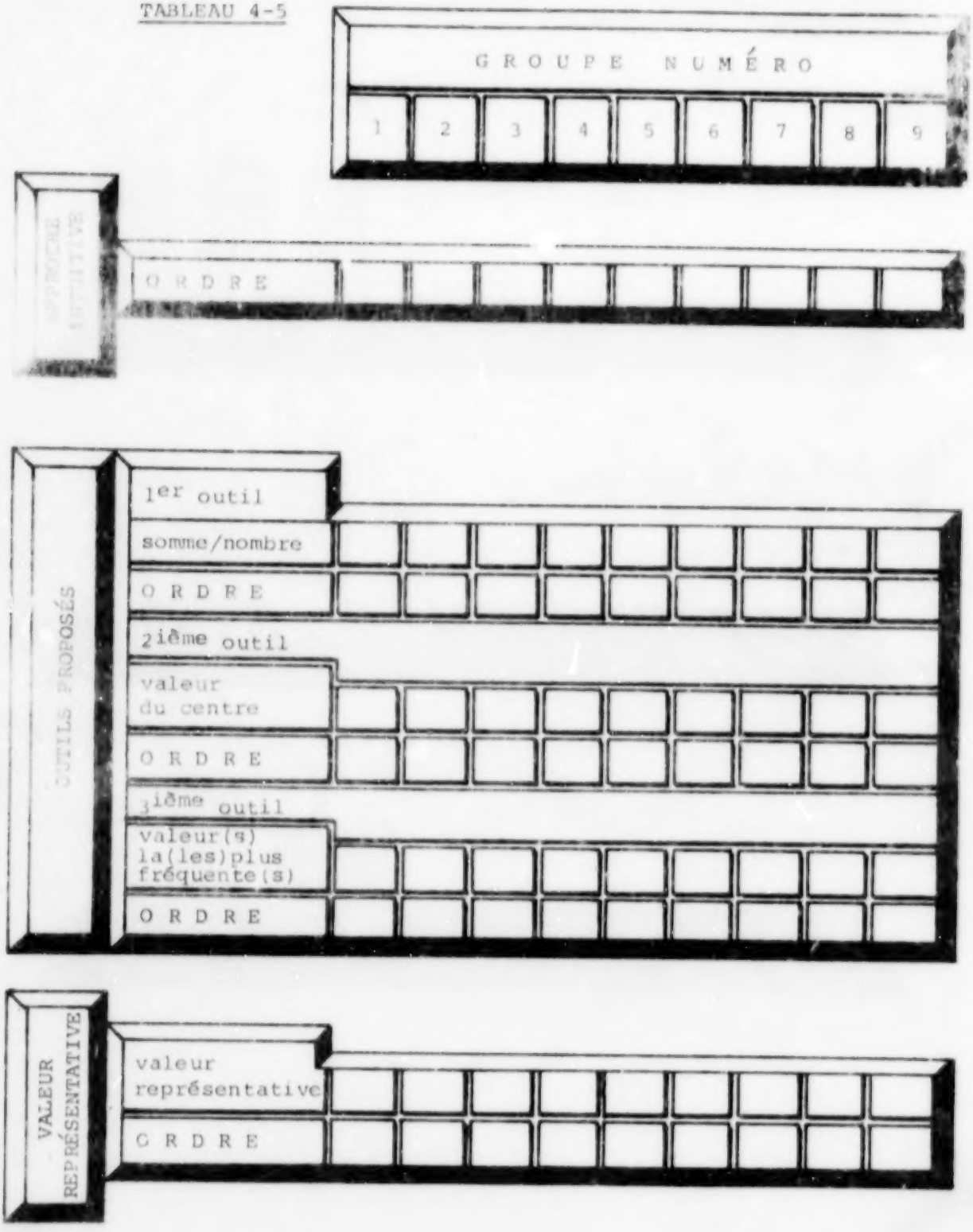
Choisir une valeur représentative de la taille de chaque groupe (répondre au tableau 4-5).

Indiquer ici les critères employés pour choisir ces valeurs.

Ordonner les groupes du plus petit au plus grand d'après les valeurs représentatives que vous venez de trouver (répondre au tableau 4-5).

Tu viens de comparer les groupes suivant leur tendance centrale.

TABLEAU 4-5


GROUPE NUMÉRO
1
2
3
4
5
6
7
8
9
APPROCHE INITIATIVE
ORDRE
OUTILS PROPOSÉS
1er outil
somme/nombre
ORDRE
2ième outil
valeur du centre
ORDRE
3ième outil
valeur(s) la(les)plus fréquente(s)
ORDRE
VALEUR REPRÉSENTATIVE
valeur représentative
ORDRE

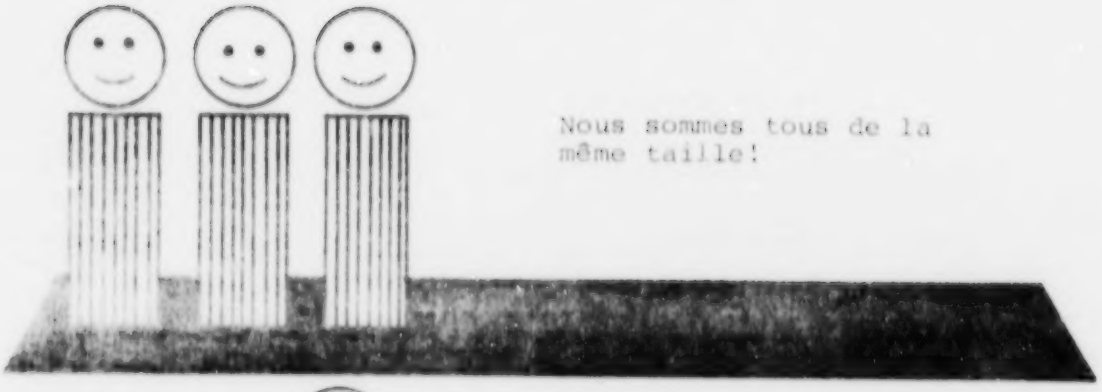
CONCLUSION

Rappeler le cheminement de l'expérience.

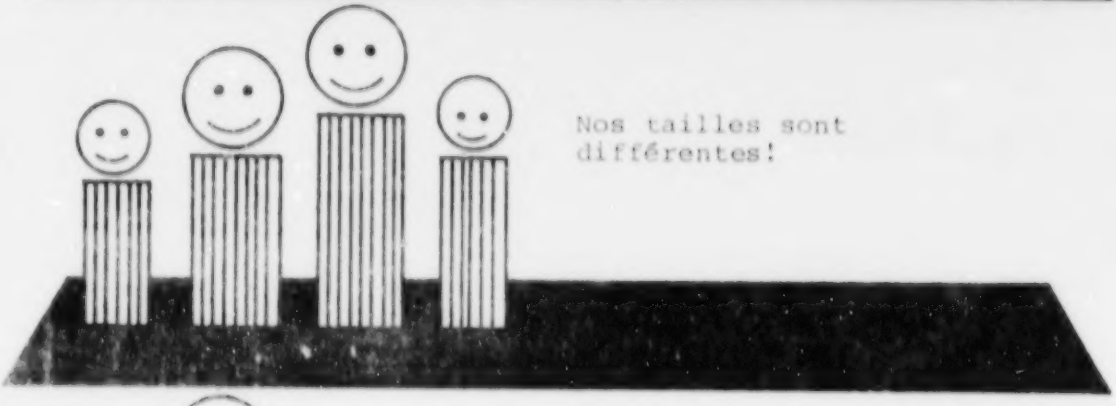
Mentionner les principales constatations.

Ton appréciation:

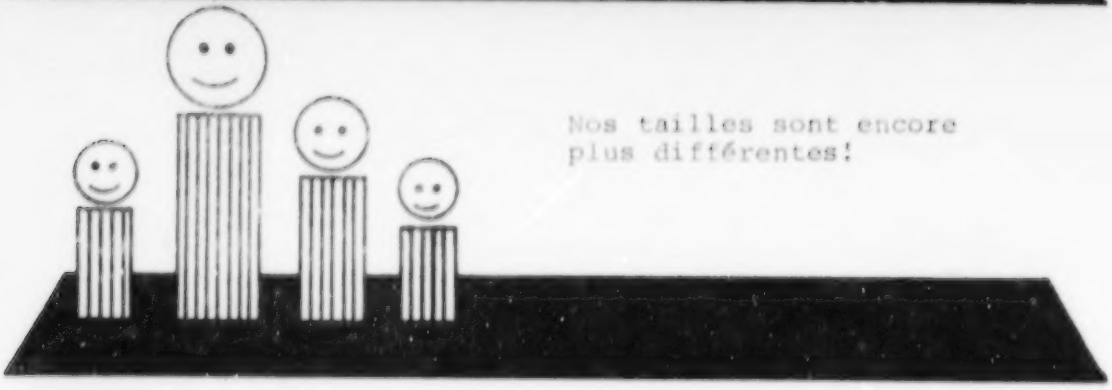
Dispersion



Nous sommes tous de la même taille!



Nos tailles sont différentes!



Nos tailles sont encore plus différentes!

EXPÉRIENCE 5

Nom(s): _____

_____

OBJECTIF GÉNÉRAL

Se sensibiliser à la dispersion et à sa mesure.

MATÉRIEL

Neuf groupes de bonshommes sur une plaque, une règle.

CONTEXTE DE L'EXPÉRIENCE

L'expérience 4 vous a donné l'occasion de comparer les tailles des bonshommes d'un groupe à l'autre; vous avez effectué ces comparaisons par le biais de la tendance centrale.

Cependant, il y a dans les groupes une autre réalité tout aussi importante du fait que les individus n'ont habituellement pas tous la même taille. Dans cette expérience vous allez tenir compte de la variation des données dans chaque groupe.

CHEMINEMENT DE L'EXPÉRIENCE

Partie A: Approche intuitive

Identifier le groupe où les tailles des bonshommes ne sont pas dispersées du tout. ☐

Justifier votre choix.

Répondre en choisissant le symbole approprié: $<$, $>$, $=$.

Q 1. *La dispersion des tailles des bonshommes du groupe 9 est _____ la dispersion des tailles des bonshommes du groupe 1.*

Q 2. *La dispersion des tailles des bonshommes du groupe 1 est _____ la dispersion des tailles des bonshommes du groupe 3.*

Q 3. *La dispersion des tailles des bonshommes du groupe 3 est _____ la dispersion des tailles des bonshommes du groupe 6.*

Q 4. *La dispersion des tailles des bonshommes du groupe 1 est _____ la dispersion des tailles des bonshommes du groupe 7.*

Classer les groupes du moins dispersé au plus dispersé (répondre dans l'espace approprié au tableau 5-3).

Indiquer les critères employés pour faire ce classement.

Partie B: Outils proposés

Un premier outil utilisable pour évaluer la dispersion de la taille pourrait consister à prendre la différence entre la plus grande et la plus petite valeur observée. (Vous pouvez vous servir des mesures déjà faites à l'expérience 4).

Calculer cette différence correspondante au tableau 5-1.

TABLEAU 5-1

Groupe numéro	1	2	3	4	5	6	7	8	9
Différence									

Un deuxième outil utilisable pour évaluer la dispersion de la taille dans un groupe pourrait consister à réaliser les étapes suivantes:

Mesurer la taille des bonshommes (vous pouvez faire ces mesures ou aller chercher les mesures déjà faites dans le cadre de l'expérience 4, tableau 4-1) (répondre au tableau 5-2).

Trouver la tendance centrale de chaque groupe (répondre au tableau 5-2).

Faire le calcul de l'écart (différence en valeur absolue) entre la taille de chaque bonhomme et la tendance centrale de la taille dans son groupe (répondre au tableau 5-2).

Additionner les écarts pour chaque groupe (répondre au tableau 5-2).

Diviser cette somme d'écarts par le nombre de bonshommes du groupe (répondre au tableau 5-2 dans la colonne intitulée moyenne des écarts).

TABLEAU 5-2

GROUPE	bonhomme	taille	tendance centrale	écart	somme des écarts	moyenne des écarts
numéro 1	1					
	2					
	3					
numéro 2	1					
	2					
	3					
	4					
	5					
	6					
numéro 3	1					
	2					
	3					
numéro 4	1					
	2					
	3					
	4					
	5					
	6					
numéro 5	1					
	2					
	3					
	4					
	5					
numéro 6	1					
	2					
	3					
numéro 7	1					
	2					
	3					
	4					
numéro 8	1					
	2					
	3					
	4					
	5					
	6					
numéro 9	1					
	2					
	3					

Partie C: Synthèse

Reporter au tableau 5-3 les valeurs obtenues avec les deux outils.

Ordonner à l'aide de chaque outil les groupes du moins disperé au plus disperé (répondre au tableau 5-3).

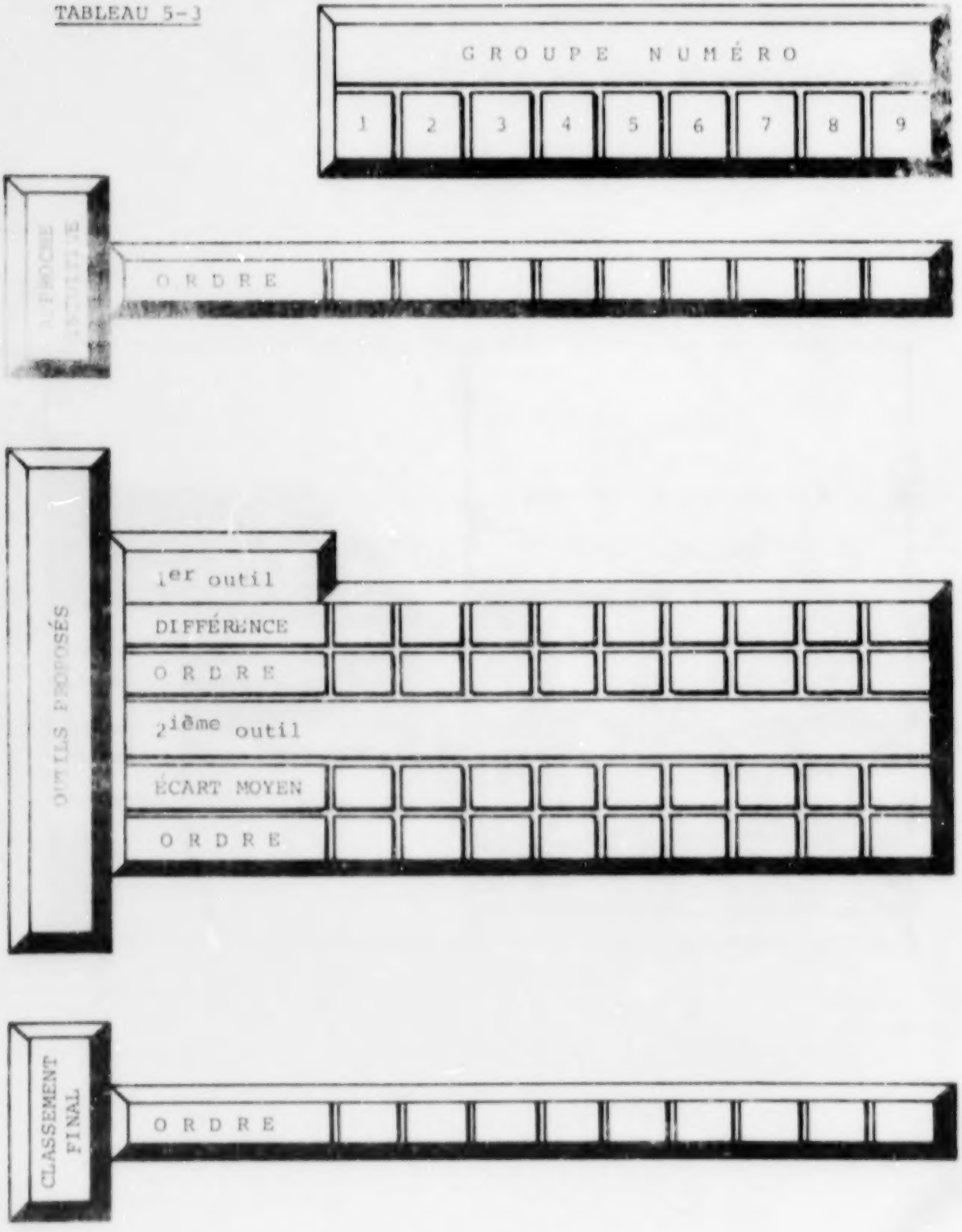
Tu peux maintenant examiner la correspondance entre l'ordre intuitif et l'ordre obtenu avec chacun des outils.

Choisir un ordre que tu accepterais finalement (répondre au tableau 5-3).

Indiquer les critères employés pour faire ce choix.

Tu viens de comparer les groupes suivant leur dispersion.

TABLEAU 5-3


GROUPE NUMÉRO
1
2
3
4
5
6
7
8
9
ORDRE
OUTILS PROPOSÉS
1er outil
DIFFÉRENCE
ORDRE
2ième outil
ÉCART MOYEN
ORDRE
CLASSEMENT FINAL
ORDRE

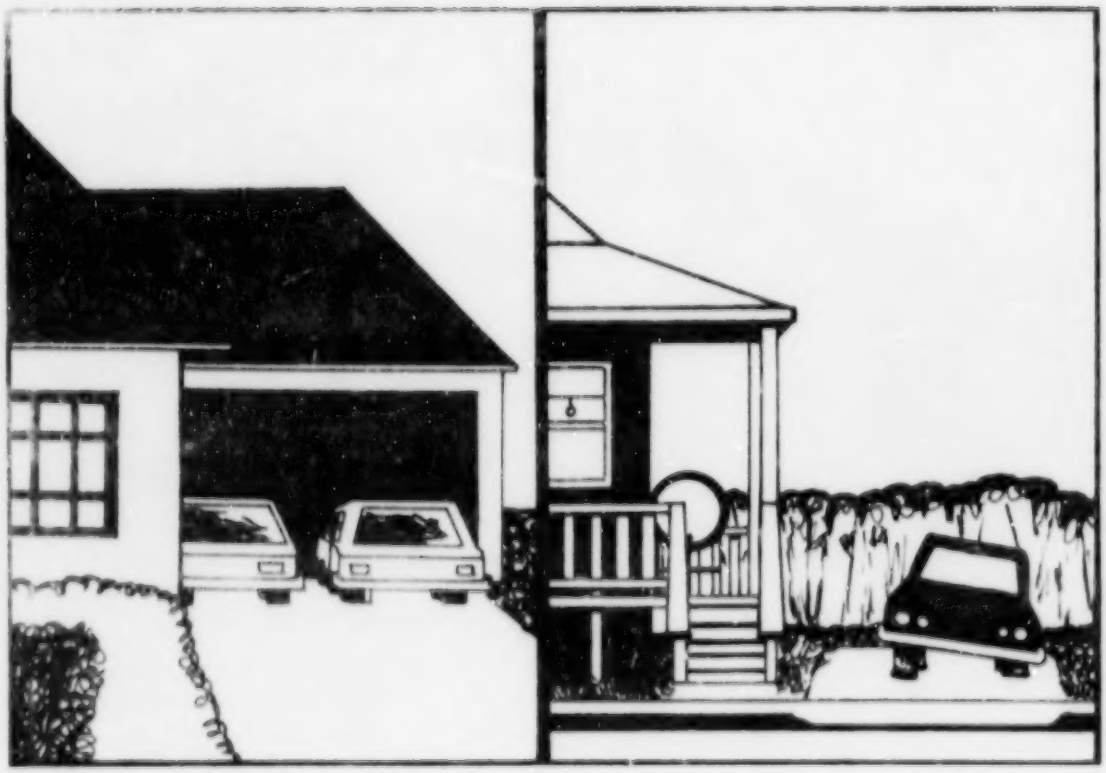
CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Corrélation



EXPÉRIENCE 6

Nom(s) _____

_____

BJECTIF GÉNÉRAL

Se sensibiliser à l'existence de liens entre deux caractéristiques d'une population et apprendre à l'exprimer graphiquement.

ATÉRIEL

Cinq groupes de bonshommes, une règle.

ONTEXTE DE L'EXPÉRIENCE

Aux expériences précédentes, nous avons appris comment décrire une population à l'aide d'une caractéristique à la fois.

Dans cette expérience, nous allons décrire une population en analysant deux caractéristiques à la fois, en cherchant principalement s'il existe un lien entre les deux caractéristiques étudiées, la nature et le degré de ce lien.

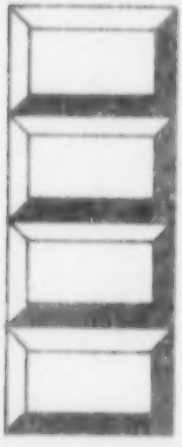
CHEMINEMENT DE L'EXPÉRIENCE

Nous allons étudier la longueur de la ceinture des bonshommes et leur taille (grandeur) pour voir s'il existe une relation entre les deux.

Au groupe 1, il existe une certaine relation entre les deux caractéristiques; en effet les plus grands bonshommes possèdent les plus longues ceintures et les plus petits les ceintures les plus courtes.

Q 1. Existe-t-il d'après toi une relation entre la taille (grandeur) des bonshommes et la longueur de leur ceinture pour les autres groupes?

Groupe	Oui ou non	Explication
2		
3		
4		
5		



Entre deux groupes d'individus, il peut exister des liens affectifs (ou relations affectives); ces liens s'ils existent peuvent être entre autres d'amour ou de haine (genre) et plus ou moins prononcés (intensité, degré ou force).

Q 2. Les relations sont-elles de même genre (et elles existent naturellement) pour les groupes?

Groupes (numéros)	Oui ou non	Explication
1 2		
1 3		
1 4		
1 5		
2 3		
2 4		
2 5		
3 4		
3 5		
4 5		

Q 8. Les relations sont-elles de même force, intensité (si elles existent naturellement) pour les groupes?

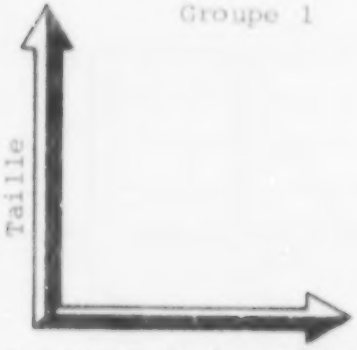
Groupes (numéros)	Oui ou non	Explication
1 – 2		
1 – 3		
1 – 4		
1 – 5		
2 – 3		
2 – 4		
2 – 5		
3 – 4		
3 – 5		
4 – 5		

Comme il est difficile de répondre à certaines questions, nous allons introduire deux méthodes pour vous aider.

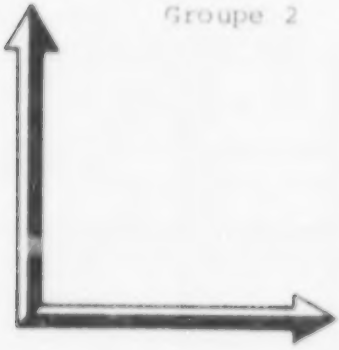
1. La méthode graphique

Faire le graphique des deux caractéristiques (taille des bonshommes et longueur de leur ceinture) pour chaque groupe.

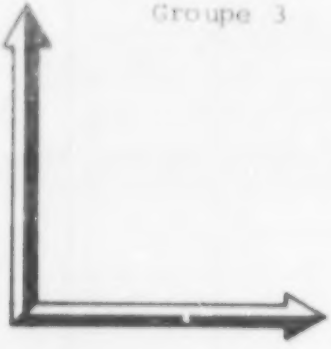
Groupe 1



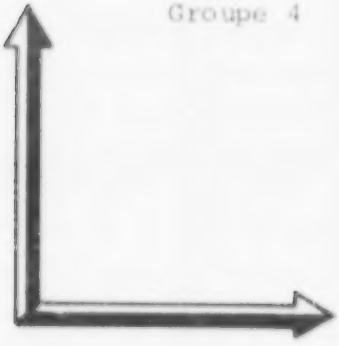
Groupe 2



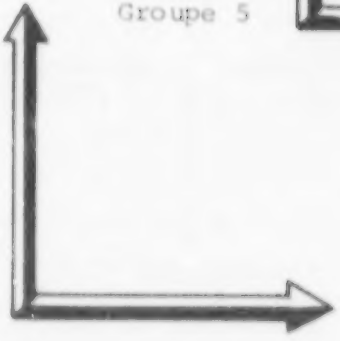
Groupe 3



Groupe 4



Groupe 5



Indiquer comment se traduit graphiquement l'existence, le genre et la force d'une relation.

2. La méthode analytique

La méthode analytique consiste à situer toutes les valeurs des deux caractéristiques par rapport à leur tendance centrale et à vérifier si elles suivent la même tendance pour la plupart des bonshommes; ceci revient à effectuer six opérations:

1. Mesurer la taille et la longueur de la ceinture de chaque bonhomme.

2. Calculer la taille moyenne et la longueur moyenne de la ceinture de tous les bonshommes.

3. Calculer pour chaque bonhomme la différence entre la taille d'un bonhomme et la taille moyenne, de même pour la longueur de la ceinture.

4. Multiplier ces différences entre elles pour chaque bonhomme.

5. Additionner tous ces produits de différence

6. Finalement diviser le tout par le nombre de bonshommes du groupe.

Effectuer ces six opérations pour chaque groupe de bonshommes et indiquer les résultats dans les tableaux 10.1 et 10.2.

TABLEAU 10-1

Groupe numéro	Taille	Longueur de la ceinture	Taille-taille moyenne (a)	Longueur-longueur moyenne de la ceinture (b)	a x b
1					
2					
3					
4					
5					

TABLEAU 10-2

Résultats	Groupe numéro				
	1	2	3	4	5
Sixième opération					

Indiquer comment se traduit par la méthode analytique l'existence, le genre et la force d'une relation (se servir des résultats du tableau 10-2).

Q 4. *Les deux méthodes sont-elles efficaces ou non pour décrire la relation entre deux caractéristiques? Quels sont les avantages et les inconvénients de chaque méthode?*

Les manufacturiers de linge savent intuitivement comme nous d'ailleurs, qu'il existe un lien entre la longueur des ceintures et la taille (grandeur) des personnes mais ils doivent connaître réellement ce lien pour confectionner des robes, des habits, des jeans, etc.; en effet quelqu'un(e) de telle taille a quelle longueur de ceinture en général ou vice versa.

Exprimer algébriquement (à l'aide d'une équation) ce lien pour le groupe 1 et le groupe 2.

1. Groupe 1

2. Groupe 2

CONCLUSION

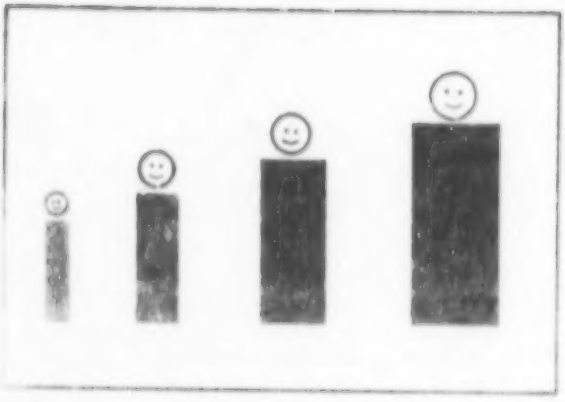
Rappeler le cheminement de l'expérience.

Mentionner les principales annotations.

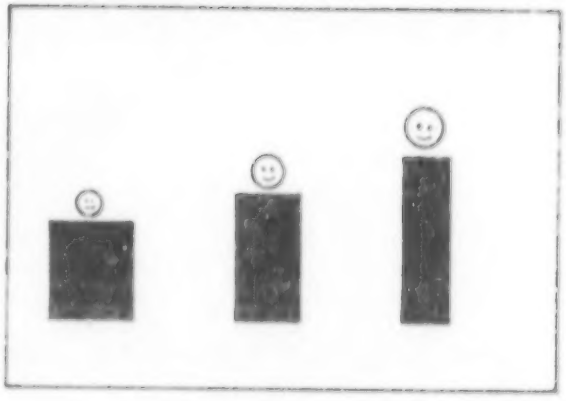
Ton appréciation:

ANNEXE

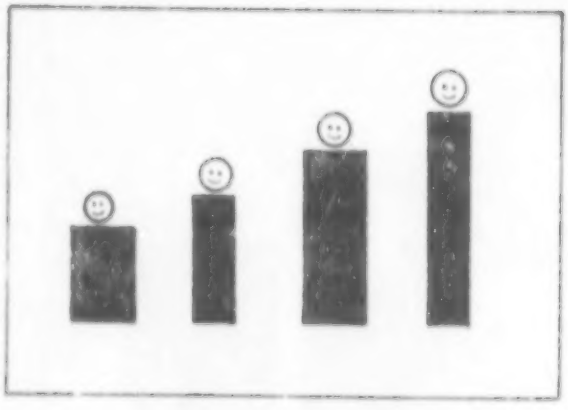
Groupe 1



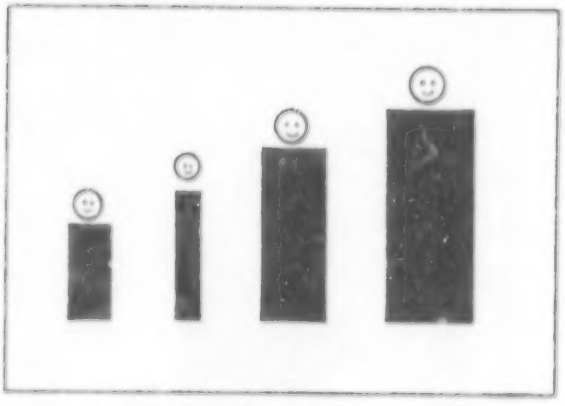
Groupe 2



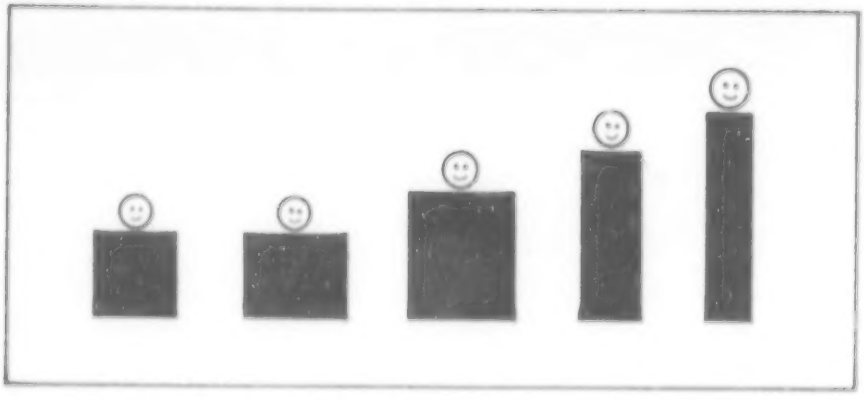
Groupe 3



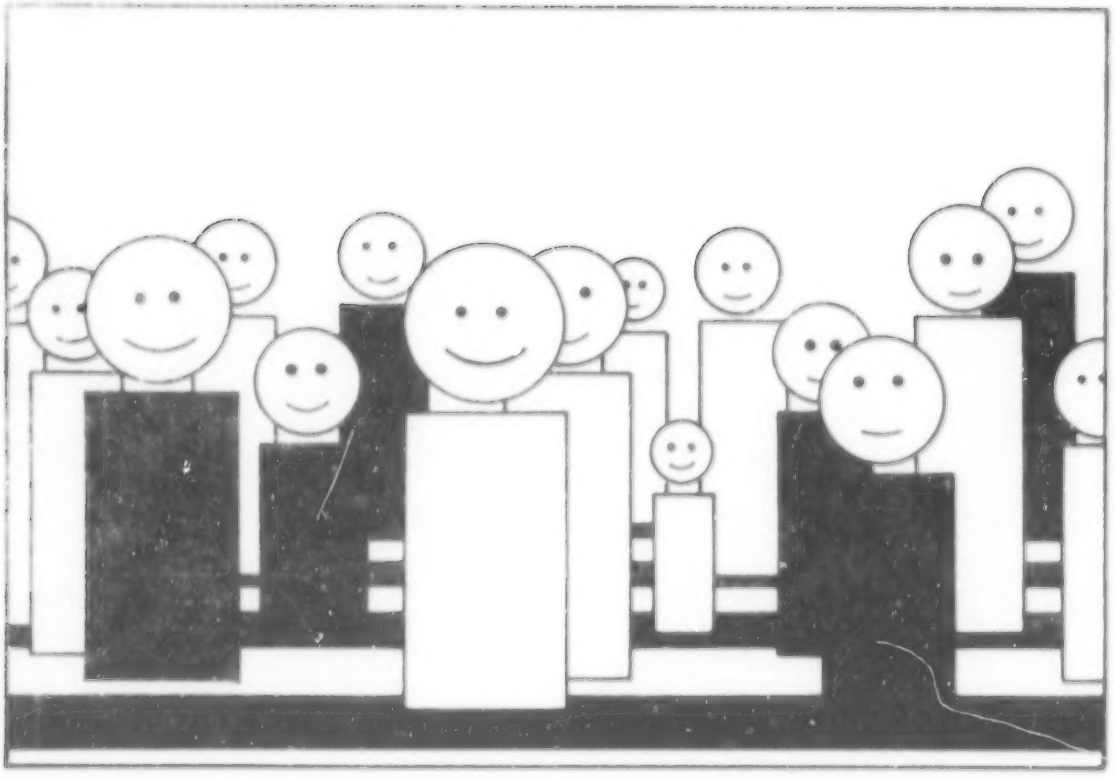
Groupe 4



Groupe 5



Population
et échantillon



EXPÉRIENCE 7

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Se sensibiliser au besoin d'échantillonner et aux méthodes d'échantillonnage.

MATÉRIEL

Un annuaire téléphonique.

CONTEXTE DE L'EXPÉRIENCE

Dans l'expérience actuelle, comme dans les expériences précédentes, vous allez faire la description d'un groupe de données. Vous constaterez qu'il n'est pas toujours nécessaire de connaître toutes les données d'un groupe pour en faire une bonne description.

CHEMINEMENT DE L'EXPÉRIENCE

On veut connaître certains renseignements au sujet de la région délimitée par l'annuaire téléphonique.

Pour répondre aux trois premières questions, tu dois te référer à cet annuaire; tu dois également indiquer ta façon de procéder (méthode).

QUESTIONS:

Q 1. Quel pourcentage de gens de sexe féminin retrouvee-tu à la page 116?

R: _____

Méthode:

Q 2. Quel est le pourcentage de gens de sexe féminin, masculin, impossible à déterminer, dans l'annuaire téléphonique de la ville où se situe ton institution, tout au moins une valeur approximative?

% masculin: _____

% féminin: _____

% impossible à déterminer: _____

Méthode:

Q 3. Quel est le pourcentage de noms francophones, anglophones, autres et impossibles à déterminer dans l'annuaire téléphonique?

% noms francophones: _____

% noms anglophones: _____

% autres noms: _____

% impossible à déterminer: _____

Méthode:

Q 4. *Comment pourrais-tu, à l'aide de l'annuaire téléphonique, trouver le nombre moyen de personnes par famille? (Tu ne dois pas te limiter à l'annuaire téléphonique).*

Q 5. *De quelle façon procèderais-tu pour évaluer le pourcentage de gens ayant le téléphone?*

CONCLUSION

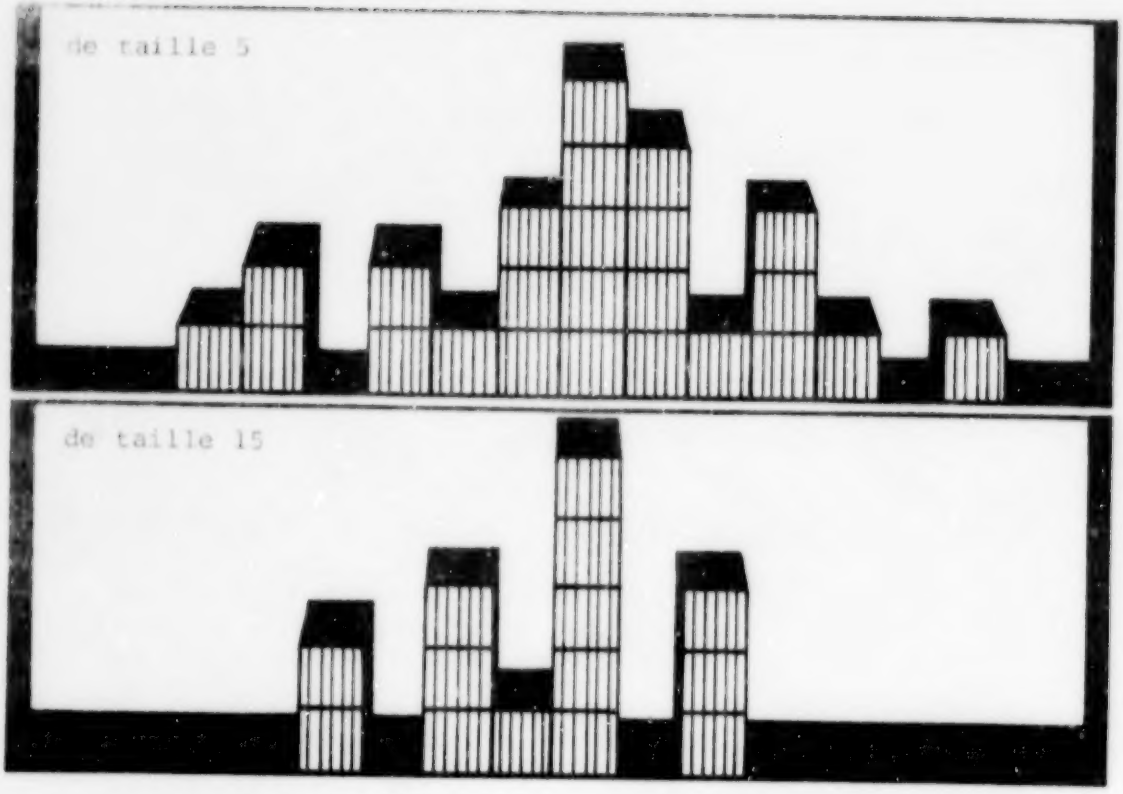
Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Échantillonnage

Distribution des moyennes d'échantillon



EXPÉRIENCE 8

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Découvrir les liens entre une population et un échantillon.

MATÉRIEL

Une règle, 300 bonshommes dans un coffre, deux tableaux collectifs, un tableau de groupe, une lisière de nombres aléatoires, des goujons.

CONTEXTE DE L'EXPÉRIENCE

Dans l'expérience 7, vous avez pu constater qu'il est souvent nécessaire de prendre un échantillon et qu'il existe différentes façons de choisir cet échantillon.

L'expérience actuelle avec une population restreinte de bonshommes nous permettra de dégager les liens entre les échantillons de différentes tailles et la population. Tout cela dans le but de vérifier jusqu'à quel point l'information recherchée au niveau de la population, se retrouve dans l'échantillon.

CHEMINEMENT DE L'EXPÉRIENCE

1. Estimation de la tendance centrale de la population.

   À cause de problèmes techniques dus aux tailles différentes des bonshommes, qui donnent une plus grande chance d'être pigés aux plus grands, tu devras prendre un moyen détourné pour le choix de ton échantillon.

Choisir une lisière de 5 et 15 nombres aléatoires.

Indiquer le numéro de ta lisière _____

Chaque nombre identifie un des 300 bonshommes du coffre.

Identifier avec l'annexe I la taille de chacun des 5 bonshommes choisis:

	1	2	3	4	5
NOMBRE ALÉATOIRE					
TAILLE					

Aller chercher dans le coffre les cinq bonshommes pour simuler ton échantillon.

Calculer la taille moyenne des 5 bonshommes de l'échantillon:

Taille moyenne: _____

? 1. Cette taille moyenne est-elle d'après toi égale à la taille moyenne des bonshommes de la population?

_____

Justifier

Se rendre à l'armoire des goujons.

Prendre 2 goujons de longueur égale à la taille moyenne des bonshommes de ton échantillon.

Déposer 1 goujon dans le tableau de groupe et un autre dans le tableau collectif.

Remettre les bonshommes dans le coffre.

Remplir le tableau en annexe II, lorsque tous les étudiants(es) auront déposé leur goujon.

Encadrer plus particulièrement ta moyenne.

Q 2. Ton opinion concernant la taille moyenne des bonshommes de la population s'est-elle modifiée, suite à la connaissance des résultats des autres étudiants(es)?

Justifier

Il serait temps maintenant de prendre connaissance du contenu du coffre, afin de vérifier la correspondance entre la taille moyenne obtenue dans ton échantillon et la taille moyenne des bonshommes de la population.

Se rendre au coffre ouvert.

Q 3. Quelle est la tendance centrale de la taille des bonshommes de la population?

_____

Tendance centrale de la taille des bonshommes de la population		Tendance centrale de la taille des bonshommes de ton échantillon
	?≃	

Q 4. La tendance centrale de la taille des bonshommes de la population est-elle à peu près égale à la tendance centrale de la taille des bonshommes de ton échantillon?

_____

Q 5. Prévois-tu avoir une meilleure estimation de la taille moyenne des bonshommes de la population, si tu prends un échantillon de 15 bonshommes?

_____

2. Liens entre différents échantillons et la population.

Revenir à ta lisière de 5 et 15 nombres aléatoires.

Identifier avec l'annexe I la taille de chacun des 15 bonshommes choisis:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NOMBRE ALÉATOIRE															
TAILLE															

Calculer à trois chiffres significatifs la taille moyenne des 15 bonshommes de l'échantillon (i.e. une décimale).

_____

Q 8. Cette nouvelle moyenne est-elle plus près de la moyenne des bonshommes de la population que celle trouvée avec un échantillon de 5 bonshommes?

_____

Se rendre à l'armoire à goujons.

Prendre 2 goujons de la longueur la plus près de la taille moyenne des bonshommes de ton échantillon.

Déposer 1 goujon dans le tableau de groupe et un autre dans le tableau collectif.

(Le tableau collectif servira ultérieurement).

PAUSE

Remplir le tableau en annexe II, lorsque tous les étudiants(es) auront déposé leur goujon.

Encadrer plus particulièrement ta moyenne.

Se référer au coffre et au tableau de groupe ou aux tableaux collectifs.

Choisir le symbole approprié: $\approx$, $>$ *ou* $<$.

1. a) Tendance centrale des moyennes d'échantillon de taille 5.

   Valeur estimée ☐

   ☐

   Tendance centrale des moyennes d'échantillon de taille 15.

   Valeur estimée ☐

   b) Tendance centrale des données de la population. ☐ Tendance centrale des moyennes d'échantillon de taille 5.

   c) Tendance centrale des données de la population. ☐ Tendance centrale des moyennes d'échantillon de taille 15.

---

2. a) Dispersion des moyennes d'échantillon de taille 5. ☐ Dispersion des moyennes d'échantillon de taille 15.

   b) Dispersion des données de la population. ☐ Dispersion des moyennes d'échantillon de taille 5.

   c) Dispersion des données de la population. ☐ Dispersion des moyennes d'échantillon de taille 15.

CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

ANNEXE I

nb	taille	nb	taille	nb	taille	nb	taille	nb	taille
1	30	61	22	121	20	181	26	241	24
2	28	62	20	122	32	182	24	242	26
3	20	63	20	123	26	183	22	243	18
4	30	64	28	124	26	184	28	244	26
5	24	65	26	125	28	185	24	245	26
6	18	66	22	126	22	186	28	246	22
7	28	67	30	127	22	187	24	247	16
8	28	68	30	128	24	188	26	248	26
9	30	69	26	129	20	189	30	249	20
10	22	70	18	130	28	190	24	250	30
11	22	71	24	131	24	191	24	251	22
12	24	72	24	132	28	192	24	252	24
13	26	73	22	133	24	193	26	253	24
14	22	74	26	134	24	194	30	254	18
15	24	75	26	135	28	195	24	255	20
16	26	76	32	136	16	196	18	256	30
17	26	77	20	137	30	197	22	257	22
18	24	78	20	138	22	198	28	258	26
19	24	79	26	139	20	199	22	259	26
20	20	80	22	140	28	200	18	260	26
21	16	81	18	141	18	201	28	261	16
22	24	82	18	142	20	202	24	262	18
23	30	83	24	143	18	203	26	263	24
24	24	84	22	144	22	204	30	264	24
25	30	85	20	145	24	205	30	265	26
26	20	86	28	146	24	206	28	266	28
27	22	87	32	147	26	207	26	267	32
28	20	88	20	148	26	208	24	268	20
29	26	89	22	149	20	209	16	269	20
30	26	90	26	150	26	210	24	270	30
31	24	91	24	151	16	211	20	271	30
32	24	92	16	152	26	212	18	272	26
33	30	93	22	153	32	213	32	273	20
34	16	94	24	154	22	214	20	274	28
35	16	95	22	155	30	215	26	275	24
36	30	96	20	156	28	216	18	276	22
37	18	97	30	157	24	217	24	277	22
38	26	98	26	158	22	218	32	278	18
39	26	99	22	159	20	219	22	279	28
40	16	100	28	160	24	220	32	280	26
41	22	101	20	161	28	221	30	281	24
42	24	102	22	162	18	222	24	282	20
43	18	103	22	163	24	223	22	283	32
44	22	104	22	164	22	224	20	284	22
45	20	105	22	165	24	225	22	285	24
46	22	106	24	166	30	226	26	286	24
47	22	107	26	167	24	227	22	287	22
48	32	108	20	168	24	228	18	288	26
49	18	109	26	169	26	229	24	289	24
50	20	110	28	170	24	230	28	290	24
51	26	111	26	171	20	231	28	291	30
52	28	112	24	172	24	232	18	292	32
53	24	113	28	173	22	233	16	293	18
54	26	114	28	174	22	234	24	294	28
55	26	115	24	175	32	235	18	295	20
56	22	116	22	176	28	236	26	296	18
57	28	117	30	177	28	237	28	297	20
58	22	118	22	178	26	238	20	298	26
59	28	119	20	179	18	239	28	299	26
60	28	120	16	180	28	240	24	300	24

TABLEAUX DE GROUPE:

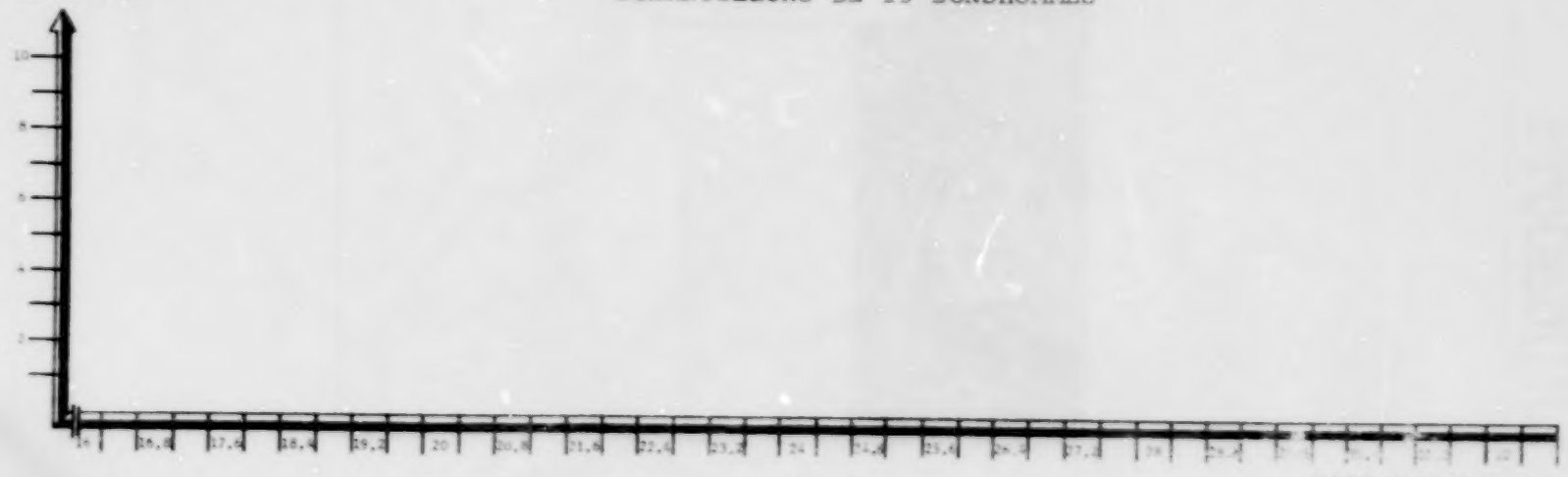
EFFECTIF

ÉCHANTILLONS DE 5 BONSHOMMES

TAILLE MOYENNE

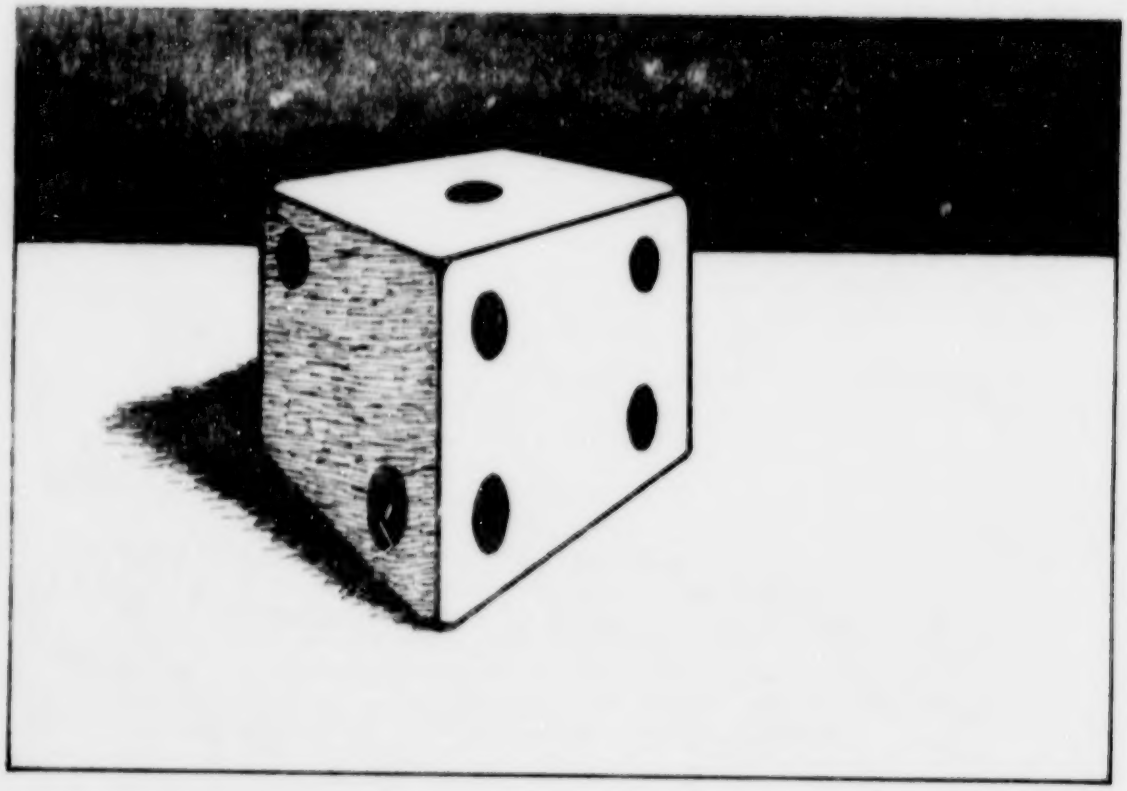
EFFECTIF

ÉCHANTILLONS DE 15 BONSHOMMES



TAILLE MOYENNE

Probabilité



EXPÉRIENCE 9

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Se sensibiliser à la notion de probabilité et à sa mesure.

MATÉRIEL

Neuf groupes de bonshommes disposés sur une plaque, une règle.

CONTEXTE DE L'EXPÉRIENCE

Dans l'expérience précédente, vous deviez en particulier évaluer la moyenne de la population; pour ce faire, vous avez utilisé des moyennes d'échantillon. Certaines de ces moyennes correspondaient à la moyenne de la population, d'autres pas. Donc avec un échantillon, il est possible de faire une erreur dans l'évaluation de la moyenne de la population.

Dans l'expérience actuelle, nous allons dégager le concept de probabilité qui nous aidera par la suite à préciser cette erreur.

CHEMINEMENT DE L'EXPÉRIENCE

Pour répondre aux questions, tu dois te référer aux neuf groupes de bonshommes du montage.

QUESTIONS:

Q 1. Si tu prends un bonhomme dans le groupe 4, quelle est la probabilité (chance) qu'il ait les yeux noirs?

Réponse: _____

Pourquoi?

Q 2. Si tu prends un bonhomme dans le groupe 5, quelle est la probabilité (chance) qu'il ait les cheveux gris?

Réponse: _____

Pourquoi?

Q 3. Y a-t-il plus de chance de retirer un bonshomme aux cheveux bruns dans le groupe 6 que dans le groupe 3?

Réponse: _____

Pourquoi?

Q 4. Quelle est la chance d'avoir un bonhomme aux cheveux bruns dans chaque groupe?

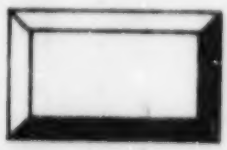
Groupe	1	2	3	4	5	6	7	8	9
Chance									

Justifier

Q 5. *Si tu prends un bonhomme dans le groupe B, quelle est la probabilité qu'il ait la (les) catégorie(s) suivante(s):*

Réponse

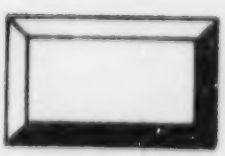
a) *les cheveux noirs ou les bras courts?*



b) *les cheveux noirs ou les bras courts ou le sexe féminin?*

c) *les bras courts ou les bras longs?*

d) *les cheveux gris ou une paire de lunettes?*

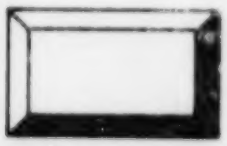


e) *la taille 32 ou les cheveux noirs?*

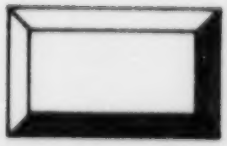
Q 6. Si tu prends un bonhomme dans le groupe 2, quelle est la probabilité qu'il ait la (les) caractéristique(s) suivante(s):

Réponse

a) la taille 22?

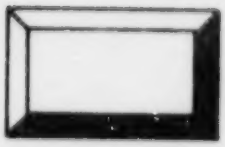


b) la taille 22 et le sexe féminin?

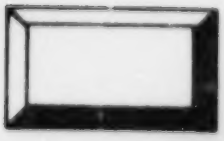


c) la taille 22 et le sexe féminin et les cheveux gris?

d) la taille 22 et le sexe féminin et les cheveux gris et les bras longs?



e) la taille 22 et le sexe féminin et les cheveux gris et les bras longs et la ceinture rouge?



Q 7. Ton(ta) voisin(e) tire un bonhomme dans le groupe 2. Évalue la probabilité que ce bonhomme ait les cheveux gris.

Réponse: _____

Q 8. Ton(ta) voisin(e) tire un bonhomme dans le groupe 2. Évalue la probabilité que ce bonhomme ait les cheveux gris s'il(s' elle) te dit qu'il a les bras longs.

Réponse: _____

Q 9. Est-ce que le fait de savoir que le bonhomme a les bras longs change la probabilité qu'il ait les cheveux gris? (Dans le groupe 2).

Réponse: _____

Justifier

Q 10. *Ton(ta) voisin(e) tire un bonhomme dans le groupe 2. Évalue la probabilité que ce bonhomme ait les bras longs.*

Réponse: _____

Q 11. *Ton(ta) voisin(e) tire un bonhomme dans le groupe 2. Évalue la probabilité que ce bonhomme ait les bras longs s'il(si elle) te dit qu'il a les cheveux gris.*

Réponse: _____

Q 12. *Est-ce que le fait de savoir que le bonhomme a les cheveux gris change la probabilité qu'il ait les bras longs? (Dans le groupe 2).*

Réponse: _____

Justifier

Q 13. *En comparant les questions 8 et 11, réponds maintenant à cette question.*

La probabilité d'avoir les cheveux gris à condition d'avoir les bras longs est-elle différente de la probabilité d'avoir les bras longs à condition d'avoir les cheveux gris?

Réponse: _____

Justifier

Q 14. *Ton(ta) voisin(e) tire un bonhomme dans le groupe 8. Est-ce que le fait de savoir que le bonhomme a les cheveux bruns change la probabilité qu'il ait les bras longs?*

Réponse: _____

Justifier

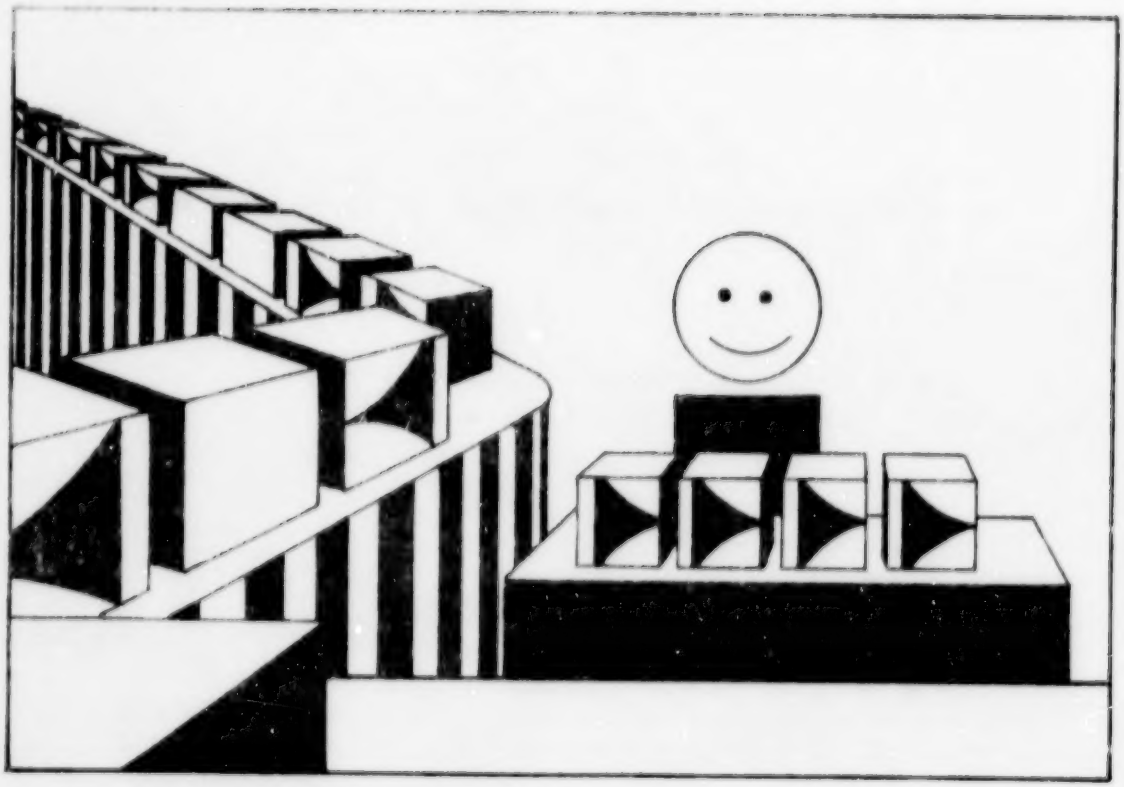
CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

Ton appréciation:

Erreur d'échantillonnage



EXPÉRIENCE 10

Nom(s) : _____

_____

OBJECTIF GÉNÉRAL

Se sensibiliser aux principaux facteurs intervenant dans l'analyse de l'erreur d'échantillonnage.

MATÉRIEL

Deux groupes de bonshommes sur une plaque, une règle.

CONTEXTE DE L'EXPÉRIENCE

À l'expérience 8, nous avons découvert les liens entre l'échantillon et la population, et les erreurs possibles si on prend un échantillon pour décrire une population. À l'expérience 9, nous avons articulé un langage qui nous permet d'exprimer les erreurs faites lors d'un échantillonnage. Dans cette expérience, nous ferons l'analyse des erreurs en termes probabilistes en vue de les contrôler lors de l'estimation de paramètres et de la vérification d'hypothèses sur la population.

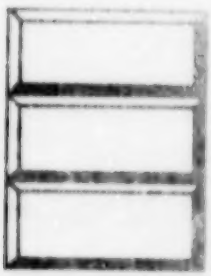
CHEMINEMENT

Q 1. *Si tu tires un bonhomme dans le groupe 1 et un bonhomme dans le groupe 7, dans quel groupe as-tu le plus de chance de tomber sur un bonhomme de:*

1) - taille 30?

2) - taille 18?

3) - taille 18?



Q 2. *Si tu tires deux bonshommes sans remise,*

a) *quels sont les échantillons possibles et différents dans le:*

i) groupe 1?

ii) groupe 7?

b) *quelle est la probabilité de tomber sur deux bonshommes dont la taille moyenne est:*

	Taille moyenne	groupe un (1)	groupe sept (7)
1.	30		
2.	18		
3.	24		
4.	28		
5.	27		
6.	29		

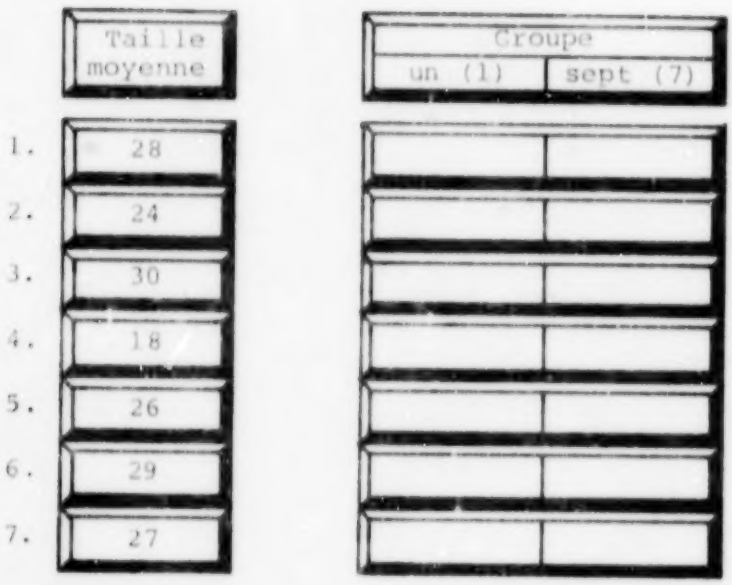
Q 3. Si tu tires deux bonshommes avec remise,

a) quels sont tous les échantillons possibles et différents dans le:

i) groupe 1?

ii) groupe 7?

b) quelle est la probabilité de tomber sur deux bonshommes dont la taille moyenne est:



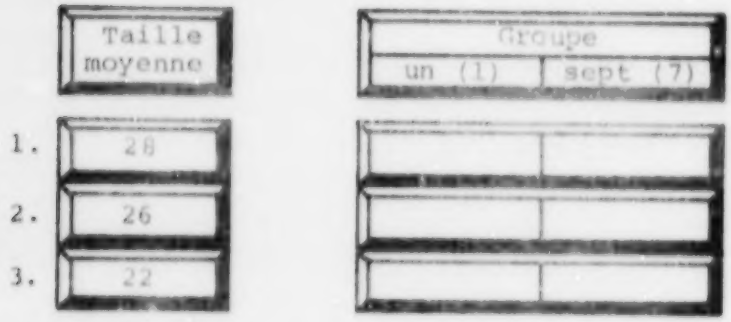
Q 4. Si tu tires trois bonshommes sans remise,

a) quels sont tous les échantillons possibles et différents dans le:

i) groupe 1?

ii) groupe 7?

b) quelle est la probabilité de tomber sur trois bonshommes dont la taille moyenne est:



Q 5. Si tu tires trois bonshommes sans remise,

a) quels sont tous les échantillons possibles et différents dans le:

i) groupe 1?

ii) groupe 7?

b) quelle est la probabilité de tomber sur trois bonshommes dont la taille moyenne est:

	Taille moyenne	Groupe un (1)	Groupe sept (7)
1.	18,0		
2.	22,0		
3.	26,0		
4.	28,0		
5.	29,3		
6.	30,0		
7.	26,6		
8.	27,3		
9.	28,6		

Q 6. Quelle est la taille moyenne des bonshommes du groupe:

1. un (1)?

2. sept (7)?

Q 7. *Si tu prends un bonhomme au hasard, quelle est la probabilité de tomber sur un bonhomme dont la taille ne s'éloigne pas plus (vers le haut ou vers le bas) de la taille moyenne de tous les bonhommes du groupe 1, ensuite du groupe 7?*

a)

Nombre d'unité(s)
1. zéro
2. une
3. deux
4. trois
5. quatre
6. six
7. huit
8. dix

b)

Groupe	
un (1)	sept (7)

Q 8. *Si tu tires sans remise deux bonhommes au hasard, quelle est la probabilité de tomber sur deux bonhommes dont la taille moyenne ne s'éloigne pas plus (vers le haut ou vers le bas) de la taille moyenne de tous les bonhommes du groupe 1, ensuite du groupe 7?*

a)

Nombre d'unité(s)
1. zéro
2. une
3. deux
4. trois
5. six
6. sept

b)

Groupe	
un (1)	sept (7)

Q 9. *Si tu tires avec remise deux bonshommes au hasard, quelle est la probabilité de tomber sur deux bonshommes dont la taille moyenne ne s'éloigne pas plus (voir a) de la taille moyenne de tous les bonshommes du groupe 1, ensuite du groupe 7?*

a)

Nombre d'unité(s)
1. zéro
2. une
3. deux
4. trois
5. six
6. sept

b)

Groupe	
un (1)	sept (7)

Q 10. *Si tu tires sans remise trois bonshommes au hasard, quelle est la probabilité de tomber sur trois bonshommes dont la taille moyenne ne s'éloigne pas plus (voir a) de la taille moyenne de tous les bonshommes du groupe 1, ensuite du groupe 7?*

a)

Nombre d'unité(s)
1. zéro
2. une
3. deux
4. trois
5. quatre
6. six
7. sept

b)

Groupe	
un (1)	sept (7)

Q 11. Si tu tires avec remise trois bonshommes au hasard, quelle est la probabilité de tomber sur trois bonshommes dont la taille moyenne ne s'éloigne pas plus (voir a) de la taille moyenne de tous les bonshommes du groupe 1, ensuite du groupe 7?

a)

Nombre d'unité(s)
1. zéro
2. une
3 deux
4. trois
5. quatre
6. six
7. sept

b)

Groupe	
un (1)	sept (7)

Q 12. Combien de bonshommes faudrait-il tirer (n = ?) sans remise pour s'assurer que l'erreur par rapport à la moyenne du groupe étudié soit inférieure ou égale à .x. unité(s) avec la probabilité .Y. dans le groupe .i.

a) un (1)

x	y	n
0	6/6	
1	6/6	

b) sept (7)

x	y	n
0	8/12	
0	24/24	

Q 13. Combien de bonshommes faudrait-il tirer (n = ?) avec remise pour s'assurer que l'erreur par rapport à la moyenne du groupe étudié soit inférieure ou égale à .x. unité(s) avec la probabilité .y. dans le groupe ...

a) un (1)

x	y	n
0	7/27	
1	7/9	
1,5	25/27	

b) sept (7)

x	y	n
0	8/16	
2	48/64	
5	8/16	

Q 14. De quel groupe (1 ou 7), l'échantillon de taille 1 ou 2 ou 3 que tu as tiré a-t-il le plus de chance de provenir si sa moyenne est ...?

Moyenne	Taille 1	Taille 2		Taille 3	
		avec remise	sans remise	avec remise	sans remise
18					
26					
28					
30					

Q 15. Si tu as à gager sur la provenance (groupe 1 ou groupe 7) de l'échantillon de taille 2 tiré avec remise, sur quel groupe miserais-tu (et indique la chance de te tromper) si la taille moyenne de l'échantillon est...?

Taille moyenne de l'échantillon tiré	Numéro du groupe misé	Probabilité d'erreur sur le choix du groupe
24		
26		
30		

CONCLUSION

Rappeler le cheminement de l'expérience.

Mentionner les principales constatations.

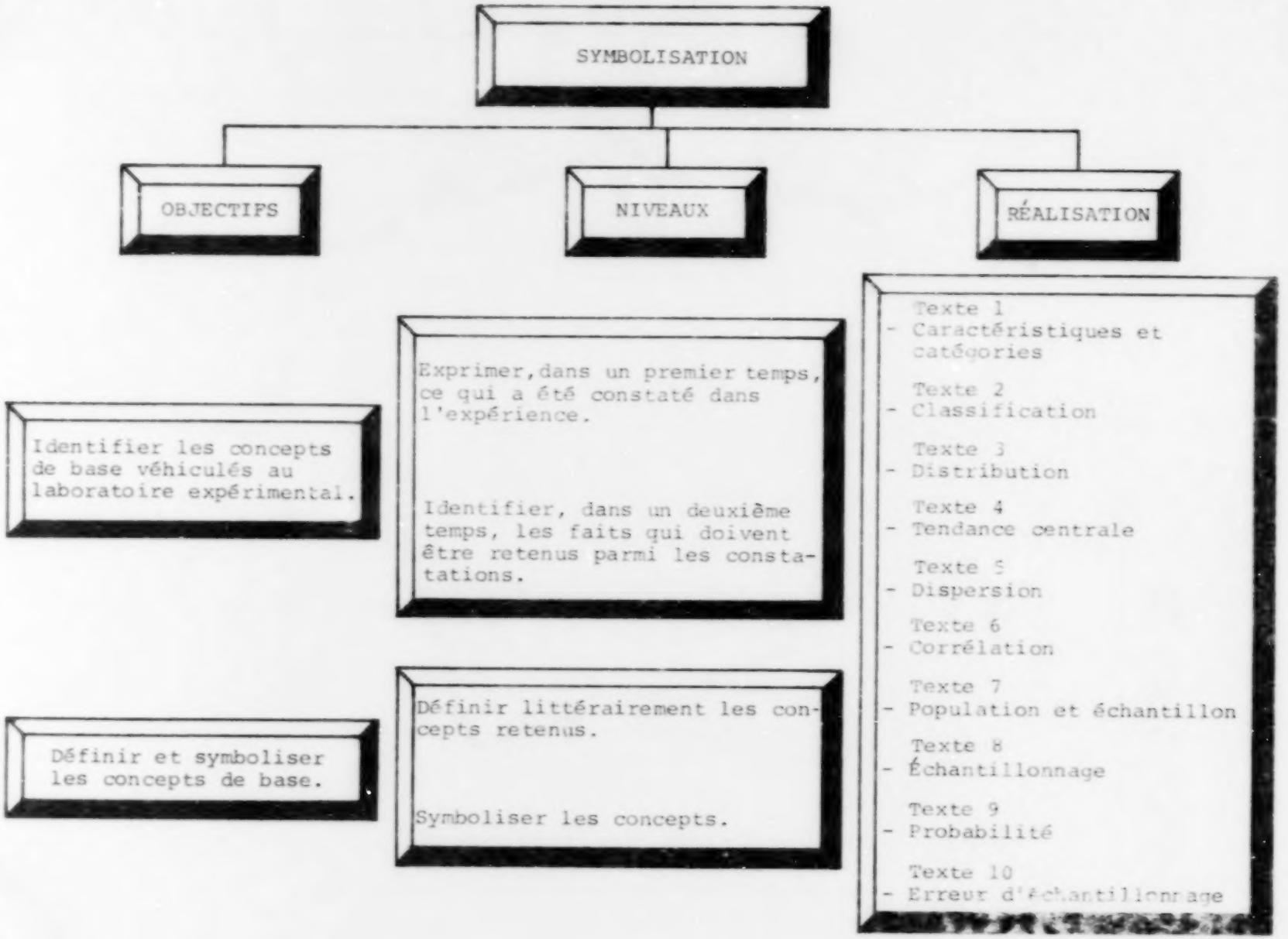
Ton appréciation:

EXERCICE

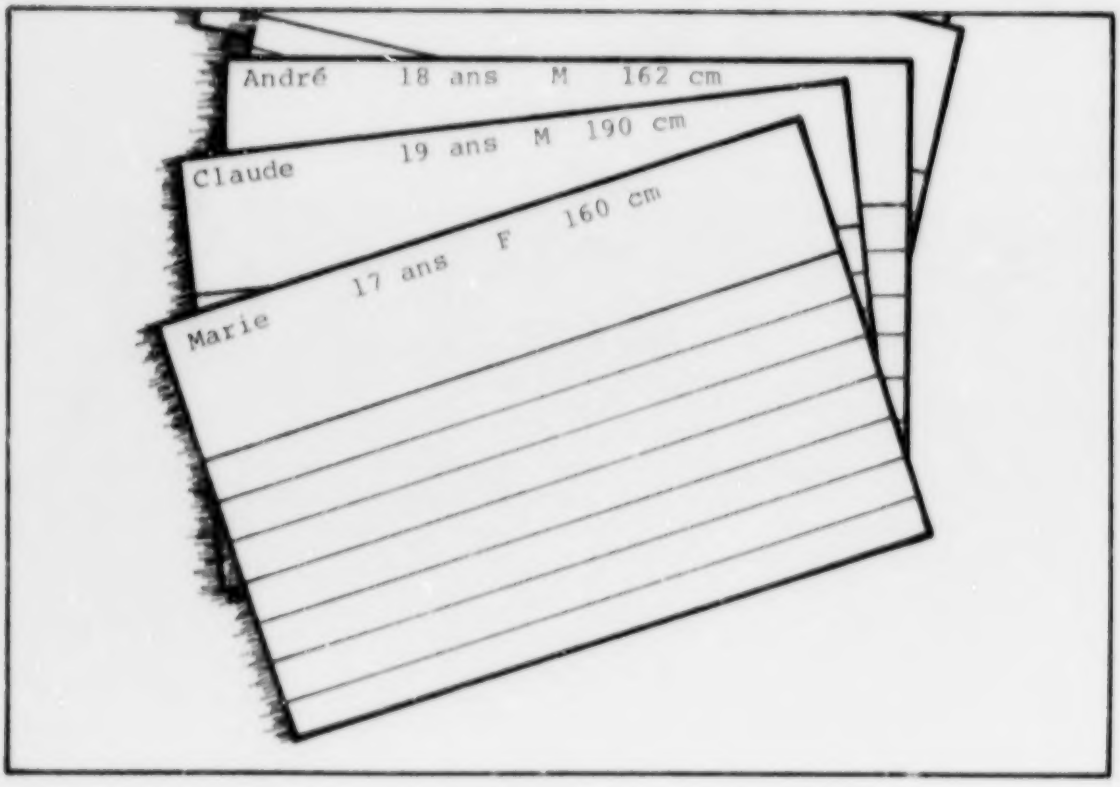
Piger 2 groupes de 10 bonshommes.

Comparer les 2 groupes de 10 bonshommes suivant leur taille en utilisant la méthode et les outils développés au laboratoire expérimental.

Symbolisation


SYMBOLISATION
OBJECTIFS
NIVEAUX
RÉALISATION
Identifier les concepts de base véhiculés au laboratoire expérimental.
Exprimer, dans un premier temps, ce qui a été constaté dans l'expérience.
Identifier, dans un deuxième temps, les faits qui doivent être retenus parmi les constatations.
Définir et symboliser les concepts de base.
Définir littérairement les concepts retenus.
Symboliser les concepts.
Texte 1
- Caractéristiques et catégories
Texte 2
- Classification
Texte 3
- Distribution
Texte 4
- Tendance centrale
Texte 5
- Dispersion
Texte 6
- Corrélation
Texte 7
- Population et échantillon
Texte 8
- Échantillonnage
Texte 9
- Probabilité
Texte 10
- Erreur d'échantillonnage

Caractéristiques et catégories



CONSTATATI

C 1.1 Plusieurs caractéristiques sont pertinentes à la description d'un groupe de bonshommes. Par exemple, la couleur des cheveux, le sexe, la taille, la couleur des vêtements, etc. ...

C 1.2 Chaque caractéristique donne lieu à diverses catégories selon les bonshommes. Par exemple, masculin et féminin pour le sexe; brun, noir et gris pour la couleur des cheveux; jeune, adolescent, adulte et vieux pour l'âge, etc. ...

C 1.3 Certaines caractéristiques sont mesurables mais d'autres ne le sont pas. Ainsi la taille (grandeur) est mesurable mais la couleur des cheveux ne l'est pas.

C 1.4 On peut regrouper toutes les données dans un tableau d'observations.

FAITS À RETENIR

F 1.1 Un tableau d'observations est un moyen de décrire un groupe.

F 1.2 Pour faire un tableau d'observations,

a) on choisit des caractéristiques;
b) on décrit chaque individu d'après les caractéristiques choisies.

F 1.3 Pour chaque caractéristique, il existe un certain nombre de catégories.

F 1.4 Seulement certaines caractéristiques sont mesurables.

DÉFINITIONS

D 1.1 **Description statistique**

Une description qui consiste à sélectionner des caractéristiques pertinentes à notre étude et à donner la catégorie correspondant à chaque caractéristique pour chaque sujet.

D 1.2 **Caractéristique (ou caractère)**

- *Sexe*
- *Couleur des cheveux*
- *Taille (grandeur)*

C'est un aspect, un trait particulier des membres d'un groupe d'individus ou d'objets.

D 1.3 **Catégorie**

- *Masculin, féminin pour le sexe*
- *Brun, gris, noir pour la couleur des cheveux*
- *20, 22, 24, ... unités pour la taille*

C'est une des subdivisions d'une caractéristique.

D 1.4 **Caractéristique quantitative (ou variable)**

- *Taille: 20, 22, ... unités*
- *Age d'un individu*

Caractéristique dont les catégories proviennent d'une mesure ou d'un dénombrement.

D 1.5 **Caractéristique qualitative**

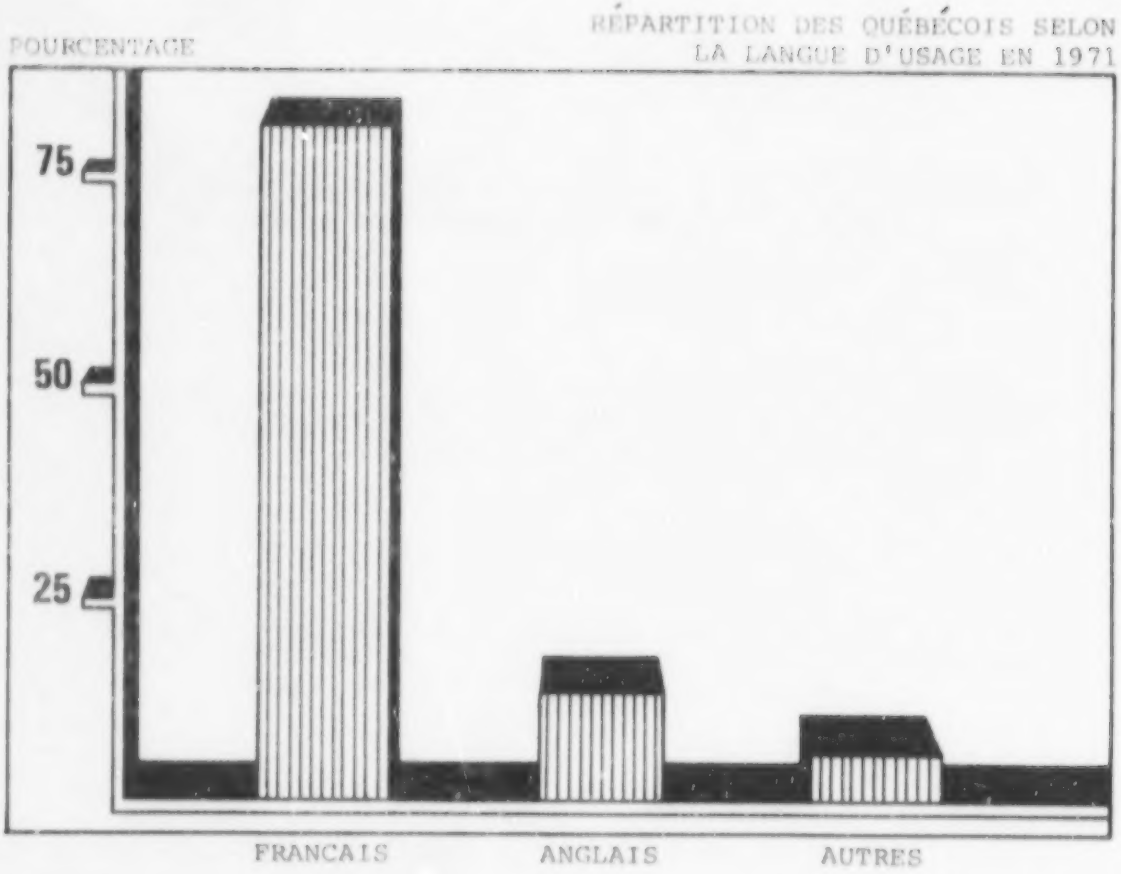
- *Sexe*
- *Taille: petit, moyen, grand*
- *Age: jeune, vieux*

Caractéristique dont les catégories ne proviennent ni d'une mesure ni d'un dénombrement.

D 1.6 Tableau d'observations

Tableau où l'on indique les catégories des individus suivant les caractéristiques retenues.

Classification



SYMBOLISATION 2

CONSTATATIONS

C 2.1 Le tableau d'effectifs de la caractéristique sexe indique le nombre de bonshommes masculins et féminins.

C 2.2 Une représentation graphique de la caractéristique sexe permet de visualiser le tableau des effectifs.

C 2.3 Les hauteurs des colonnes du graphique de la caractéristique sexe sont proportionnelles au nombre de bonshommes par file sur la plaque.

C 2.4 Les trois dernières constatations sont aussi valables en rapport avec les caractéristiques suivantes: couleur des cheveux, taille.

C 2.5 Les mêmes bonshommes se répartissent différemment selon que l'on étudie les caractéristiques sexe, couleur des cheveux ou taille.

C 2.6 Les catégories de la caractéristique taille peuvent être ordonnées, ce qui est impossible pour le sexe ou la couleur des cheveux.

FAITS À RETENIR

F 2.1 Un tableau d'effectifs est un moyen de décrire un groupe suivant un aspect particulier (caractéristique).

F 2.2 Pour construire un tableau d'effectifs,

a) on choisit une caractéristique,
b) on énumère les catégories différentes,
c) on compte le nombre d'individus appartenant à chaque catégorie.

F 2.3 Un bon moyen de visualiser la distribution d'une caractéristique est d'en faire une représentation graphique.

F 2.4 Il est avantageux d'ordonner les catégories d'une caractéristique quand c'est possible. C'est le cas de toutes les caractéristiques quantitatives et même de certaines caractéristiques qualitatives.

F 2.5 Le tableau d'effectifs fait perdre l'identité des observations.

F 2.6 La somme des effectifs égale l'effectif total.

F 2.7 Par convention, dans une représentation graphique, on indique les catégories sur l'axe horizontal et les effectifs sur l'axe vertical.

DÉFINITIONS

D 2.1 **Effectif** (e_i)

Sexe	Effectif
M	9
F	16
Total	25

- 9 est l'effectif de la catégorie masculin
- 16 est l'effectif de la catégorie féminin

C'est le nombre de sujets d'une catégorie.

D 2.2 **Effectif total**

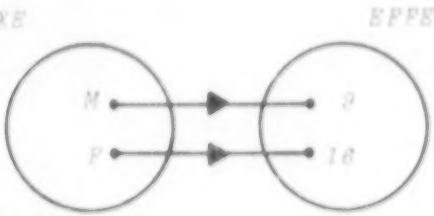
- 25 dans le tableau ci-dessus

C'est le nombre total de sujets d'un groupe.

$$\text{Effectif total} = \sum e_i$$

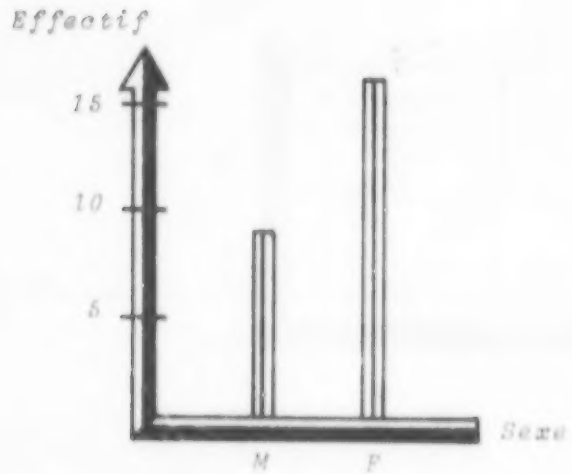
où $\sum$ représente la somme.

D 2.3 **Distribution d'une caractéristique**



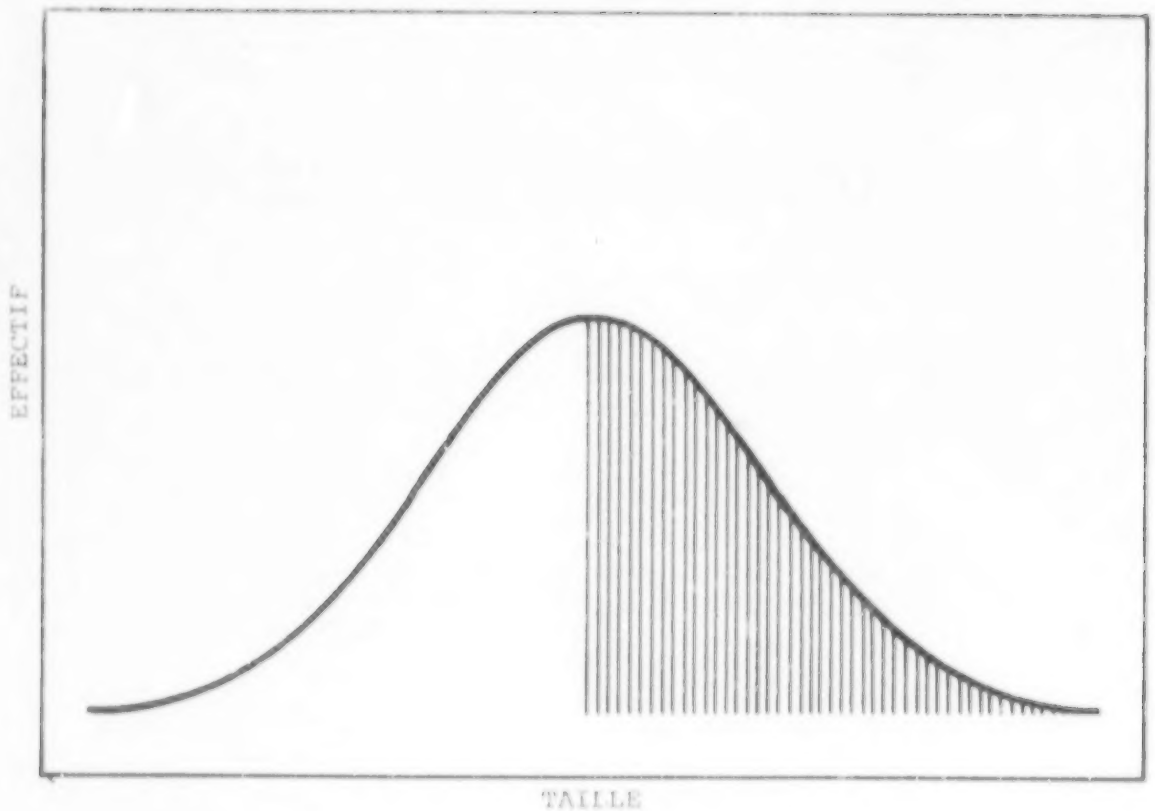
C'est la correspondance qui existe entre chaque catégorie d'une caractéristique et son effectif.

D 2.4 Représentation graphique de la distribution d'une caractéristique



C'est le dessin d'une distribution.

Distribution



CONSTATATIONS

C 3.1 Certains groupes ont des contours semblables.

C 3.2 Certains groupes ont des pics, d'autres pas.

C 3.3 Certains de ces groupes ont un pic central.

C 3.4 Certains de ces groupes ont une répartition symétrique par rapport au pic central.

C 3.5 Des groupes se ressemblent à partir d'un ou de plusieurs de ces critères.

C 3.6 On peut évaluer le pourcentage d'individus ayant une taille comprise dans un certain intervalle.

C 3.7 Même si les groupe 4 et 5 ont des formes semblables de distribution, les tailles de bonshommes sont réparties différemment.

FAITS À RETENIR

F 3.1 On peut retenir quatre critères pour reconnaître ou comparer les formes de distribution:

1) les contours,
2) la présence, la position et le nombre de pics,
3) la symétrie,
4) le pourcentage de données entre deux valeurs.

F 3.2 Deux groupes ayant le même nombre de données peuvent avoir différentes formes de distribution.

DÉFINITIONS

D 3.1 Contour

- Le groupe 1 a un contour rectangulaire.
- Le groupe 3 a un contour triangulaire.

La ligne qui encadre l'ensemble des unités de la distribution occupées par chaque bonhomme.

D 3.2 Pic

- Le groupe 3 a un pic.
- Le groupe 2 a deux pics.
- Le groupe 1 n'a pas de pic.

Prédominance en termes d'effectifs d'une catégorie par rapport à ses voisines.

D 3.3 Pic central

- Le pic du groupe 4 et celui du groupe 5 sont des pics centraux.
- Le groupe 3 a un pic qui n'est pas central.

Pic situé au centre de la distribution.

D 3.4 Symétrie

- Le groupe 4 est symétrique par rapport au pic central.
- Le groupe 1 est symétrique par rapport à la catégorie centrale

Répartition identique des bonshommes de chaque côté d'un axe de référence.

D 3.5 Fréquence (f_i)

- La fréquence de la taille 24 dans le groupe 3 est de 4/10 (0,4 ou 40%).
- La fréquence de la taille 22 dans le groupe 5 est de 4/20, (0,2 ou 20%).

- *La fréquence de chaque catégorie du groupe 1 est de 2/10 (0,2 ou 20%).*

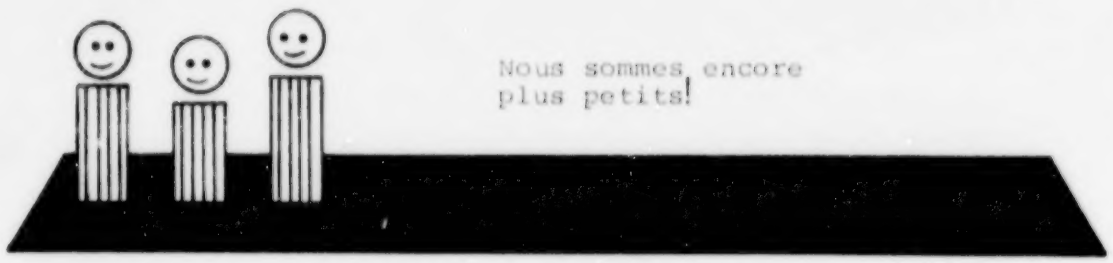
Le nombre de bonshommes d'une catégorie par rapport au nombre total de bonshommes dans le groupe.

$$\text{Fréquence} = \frac{\text{effectif de la catégorie}}{\text{effectif total}}$$

Tendance centrale

Nous sommes grands!

Nous sommes petits!



Nous sommes encore plus petits!

CONSTATATIONS

C 4.1 On constate qu'intuitivement:

a) la taille des bonshommes du groupe 1 est plus grande que la taille des bonshommes du groupe 3;

b) la taille des bonshommes du groupe 6 est plus grande que la taille des bonshommes du groupe 3;

c) la taille des bonshommes du groupe 1 est plus grande que la taille des bonshommes du groupe 2;

d) la taille des bonshommes du groupe 1 est la même que la taille des bonshommes du groupe 4;

e) il est difficile de comparer la tendance centrale du groupe 5 avec celle du groupe 8.

C 4.2 On constate que le premier outil:

a) nous donne des résultats en accord avec l'ordre intuitif établi pour les groupes 1, 2, 3, 4, 6 et 9;

b) nous donne des valeurs permettant de comparer les groupes 7 et 8 avec les autres groupes;

c) tient compte de la taille de chaque bonhomme d'un groupe;

d) ne donne pas nécessairement une valeur correspondant exactement à quelqu'un du groupe.

C 4.3 On constate que le deuxième outil:

a) donne des résultats en accord avec l'ordre intuitif dans le cas des groupes 1, 2, 3, 5, 6 et 9;

b) donne une valeur au centre des tailles d'un groupe;

c) pose des problèmes de choix d'une valeur centrale dans les groupes 7 et 8 étant donné que le centre se situe entre deux bonshommes.

C 4.4 On constate que le troisième outil:

a) donne des résultats en accord avec l'ordre intuitif dans le cas des groupes 2, 8 et 9;

b) donne deux résultats dans le cas du groupe 5;

c) ne donne pas de résultats significatifs dans les groupes 1, 3, 4 et 6;

d) donne toujours (lorsqu'il s'applique) une valeur correspondant à plusieurs observations dans le groupe.

C 4.5 On constate qu'il y a plus d'une approche possible pour évaluer la tendance centrale d'un groupe.

FAITS À RETENIR

F 4.1 Il existe différentes approches permettant d'évaluer la tendance centrale d'un groupe; un bon outil ne doit pas contredire l'intuition mais plutôt la préciser.

F 4.2 Chacune des approches comporte des avantages et des inconvénients:

Outil 1

Il tient compte de toutes les données mais le résultat peut quelquefois donner une fausse idée du groupe.

Outil 2

Il donne le centre du groupe mais ne tient pas compte de l'écart entre les données consécutives.

Outil 3

Il est habituellement facile d'emploi mais il ne tient pas compte de toutes les données et peut en plus donner des résultats multiples ou n'en donner aucun.

F 4.3 Les mesures de tendance centrale permettent de comparer des groupes.

DÉFINITIONS

D 4.1 Tendance centrale

- *La tendance centrale de la taille du groupe 1 est égale à la tendance centrale du groupe 4.*
- *La tendance centrale du groupe 1 est plus élevée que la tendance centrale du groupe 8.*

La tendance centrale est un des aspects de la description d'un groupe de données numériques visant à déterminer une valeur représentative de l'ensemble des données.

D 4.2 Outils de mesures de tendance centrale

a) la moyenne arithmétique (le premier outil) ($\bar{x}$ ou μ)

- *La somme des tailles divisée par le nombre de bonshommes du groupe 1 est égale à 28 unités.*
- *La somme des tailles divisée par le nombre de bonshommes du groupe 8 est égale à 21,6 unités.*

La moyenne arithmétique est la valeur obtenue suite à une redistribution uniforme de la somme des données.

$$\text{Moyenne arithmétique} = \frac{\text{somme des données}}{\text{nombre de données}}$$

b) la médiane (le deuxième outil) (md)

- *La valeur centrale de la taille du groupe 1 est égale à 28 unités.*
- *La valeur centrale de la taille du groupe 8 est égale à 19 unités.*

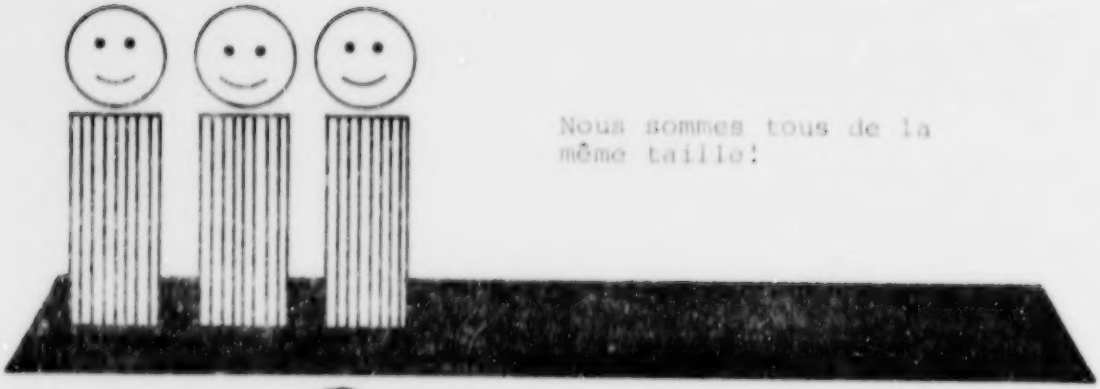
La médiane est une valeur telle qu'au moins la moitié des données ordonnées sont plus petites ou égales à cette valeur et au moins la moitié des données sont plus grandes ou égales à cette valeur.

c) le mode (le troisième outil) (mo)

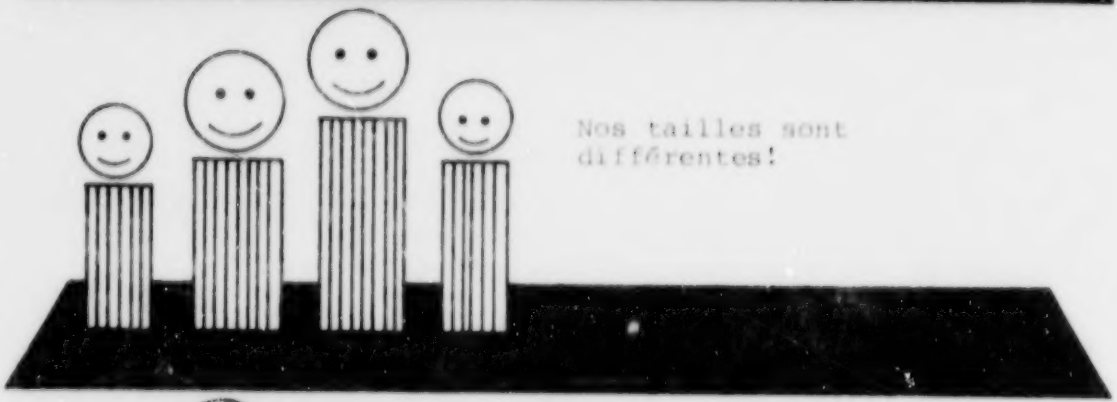
- *La valeur de la taille la plus fréquente du groupe 2 est égale à 22 unités.*
- *La valeur de la taille la plus fréquente du groupe 8 est égale à 18 unités.*

Le mode est la ou les catérorie(s) la(les) plus fréquente(s) (celle(s) qui a(ont) l'effectif le plus élevé).

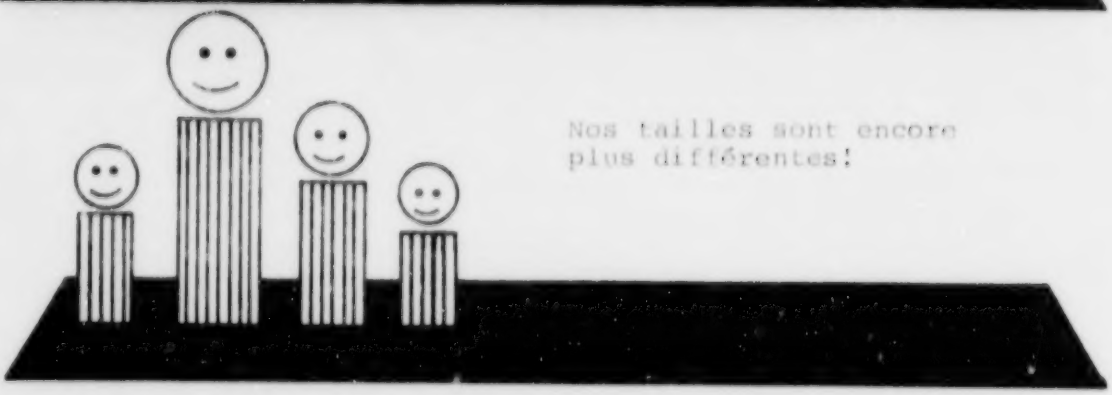
Dispersion



Nous sommes tous de la même taille!



Nos tailles sont différentes!



Nos tailles sont encore plus différentes!

CONSTATATIONS

C 5.1 On constate qu'intuitivement:

a) les bonshommes du groupe 9 sont de tailles identiques; leurs tailles ne sont pas dispersées;

b) les tailles des bonshommes du groupe 3 sont plus dispersées que celles des bonshommes du groupe 1;

c) les tailles des bonshommes du groupe 7 sont plus dispersées que celles des bonshommes du groupe 1;

d) les tailles des bonshommes du groupe 3 et du groupe 6 sont dispersées également;

e) les tailles des bonshommes du groupe 1 et du groupe 4 sont dispersées également;

f) les tailles des bonshommes du groupe 2 sont moins dispersées que celles des bonshommes du groupe 3;

h) malgré ces éléments, il demeure difficile de classer l'ensemble des groupes selon la dispersion; entre autres, les groupes 2, 5 et 8 sont difficiles à situer.

C 5.2 On constate que le premier outil:

a) donne une réponse pour tous les groupes et que ces réponses permettent de classer les groupes;

b) considère égales les dispersions des groupes 2 et 3 et ce, contrairement à notre intuition.

C 5.3 On constate que le deuxième outil:

a) donne une valeur de dispersion pour tous les groupes et que l'on peut classer les groupes selon ces valeurs;

b) fait une moyenne des écarts;

c) donne un ordre qui ressemble à ceux obtenus au moyen des deux méthodes précédentes.

C 5.4 On constate que le classement final est difficile, étant donné que chaque façon de procéder à ses mérites.

FAITS À RETENIR

F 5.1 Il existe différentes approches permettant d'évaluer la dispersion d'un groupe; un bon outil ne devrait pas contredire l'intuition mais plutôt la préciser.

F 5.2 Chaque approche a ses avantages et ses inconvénients:

Outil 1

Il est très facile et rapide à utiliser, mais ne tient compte que de deux données extrêmes; deux données fournissent à elles seules une valeur de dispersion de l'ensemble des données.

Outil 2

Il tient compte individuellement de chaque donnée, mais il comporte davantage de calculs.

F 5.3 Les mesures de dispersion permettent de comparer des groupes.

DÉFINITIONS

D 5.1 Dispersion

- *Les tailles du groupe 8 ne sont pas dispersées.*
- *Les tailles du groupe 1 sont dispersées.*

La dispersion est un des aspects de la description d'un groupe de données numériques visant à rendre compte de la variation des données dans le groupe.

D 5.2 Outils de mesure de dispersion

a) l'étendue simple (R)

- *La différence entre le plus grand et le plus petit bonhomme du groupe numéro un est égale à 4 unités.*

L'étendue simple est la différence entre la donnée la plus élevée et la donnée la plus basse d'une caractéristique d'un groupe.

Étendue simple = donnée la plus élevée - donnée la plus basse.

b) l'écart moyen (E.M.)

- *La moyenne des écarts autour de la moyenne des bonshommes du groupe numéro un égale 1,33 unités.*
- *La moyenne des écarts autour de la moyenne des bonshommes du groupe numéro huit égale 4,22 unités.*

L'écart moyen est la valeur moyenne des écarts (différences en valeur absolue) entre chaque donnée et la moyenne du groupe.

Écart moyen = $\frac{\text{somme des écarts entre chaque donnée et la moyenne}}{\text{nombre des données}}$

Corrélation

CONSTATATIONS

C 6.1 On constate qu'intuitivement:

a) pour tous les bonshommes du groupe 1, plus il est grand plus sa ceinture est longue;

b) pour tous les bonshommes du groupe 2, plus il est grand plus sa ceinture est courte;

c) pour la plupart des bonshommes du groupe 4, plus il est grand plus sa ceinture est longue;

d) pour la plupart des bonshommes du groupe 5, plus il est grand plus sa ceinture est courte;

e) pour les bonshommes du groupe 3, les grands comme les petits peuvent avoir une ceinture longue ou courte;

f) il existe une relation entre la taille des bonshommes et la longueur de leur ceinture dans les groupes 1, 2, 4 et 5 mais pas dans le groupe 3;

g) les relations dans les groupes 1 et 4 sont de même genre mais de genre opposé à celles dans les groupes 2 et 5;

h) la relation dans le groupe 4 est plus forte que celle dans le groupe 3 mais plus faible que celle dans le groupe 1;

i) la relation dans le groupe 5 est plus faible que celle dans le groupe 2;

j) il semble difficile de comparer la force des relations qui ne sont pas du même genre.

C 6.2 La méthode graphique montre que:

a) les points représentant la taille et la longueur de ceinture des bonshommes du groupe 1 sont tous alignés suivant une droite, ce qui traduit l'existence d'une relation linéaire parfaite entre ces deux caractéristiques. Le même phénomène existe dans le groupe 2. Toutefois les deux droites ont des pentes de signe opposé;

b) dans les autres groupes, les points sont moins alignés, ce qui illustre l'existence de relations moins fortes entre les deux caractéristiques;

c) il est encore difficile de comparer la force des relations dans les groupes 4 et 5.

C 6.3 Le signe de la réponse obtenue par la méthode analytique nous renseigne sur le genre de la relation qui existe entre deux caractéristiques.

C 6.4 Les valeurs obtenues par cette méthode dans les groupes 1 et 2 (2,5 par rapport à 1,33 abstraction faite du signe) nous porte à croire que la relation dans le groupe 1 est plus forte que celle dans le groupe 2. Par contre, la méthode graphique montre que ces deux relations sont de même force.

C 6.5 Dans le groupe 1, la relation entre la taille des bonshommes et la longueur de leur ceinture peut s'exprimer algébriquement par l'équation:

taille = 2 x longueur de ceinture + 5

C 6.6 Dans le groupe 2, l'expression algébrique de la relation entre les deux caractéristiques est:

taille = -2 x longueur de ceinture + 16

FAITS À RETENIR

F 6.1 Dans un groupe donné, il peut exister une relation entre deux caractéristiques; si elle existe, elle est soit parfaite ou imparfaite.

F 6.2 Une relation peut être du genre positif ou négatif selon que les valeurs fortes d'une caractéristique correspondent soit aux valeurs fortes de l'autre caractéristique soit à ses valeurs faibles.

F 6.3 a) Il existe différentes méthodes pour détecter l'existence, le genre et la force d'une relation. Une bonne méthode ne doit pas contredire l'intuition mais la préciser.

b) Les deux méthodes utilisées à l'expérience ont leurs avantages et leurs inconvénients;

Méthode graphique

Elle n'exige aucun calcul, fait ressortir généralement l'existence et le genre de la relation et permet d'identifier les relations parfaites; par contre, elle permet difficilement la comparaison de relations de force voisine, surtout lorsqu'elles sont de genre opposé.

Méthode analytique

Elle permet de trouver le genre d'une relation mais non sa force; les calculs sont longs; elle tient compte de la tendance centrale et du nombre de données mais non de leur dispersion.

F 6.4 Il est souvent utile d'exprimer algébriquement le lien qui existe entre deux caractéristiques.

DÉFINITIONS

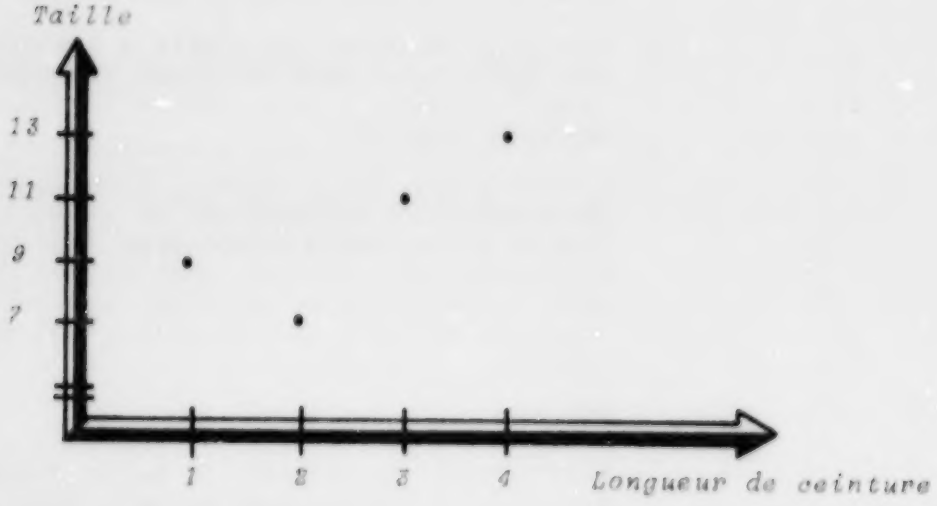
D 6.1 Corrélation

- *La relation entre la taille et la longueur de ceinture des bonshommes du groupe 1.*

La dépendance qui existe entre deux caractéristiques quantitatives.

D 6.2 Nuage de points

- *Le nuage de points du groupe 4 est:*



Graphique faisant correspondre à chaque paire de valeurs des caractéristiques un point de ce graphique.

D 6.3 Covariance (Cov)

- *La réponse obtenue après avoir effectué les six opérations sur la taille et la longueur de ceinture des bonshommes du groupe 5 est -2,3.*

Mesure de la variabilité conjointe de deux caractéristiques quantitatives autour de leurs moyennes

$$\text{Covariance} = \frac{\sum_{i=1}^{n} (x_i - \text{m.a. des } x_i)(y_i - \text{m.a. des } y_i)}{\text{nombre des } x_i \text{ ou des } y_i}$$

D 6.4 Régression

- *L'expression algébrique de la relation entre les deux caractéristiques dans le groupe 1 est:*

  taille = 2 x longueur de ceinture + 5.

Recherche d'une équation exprimant la liaison qui existe entre deux variables.

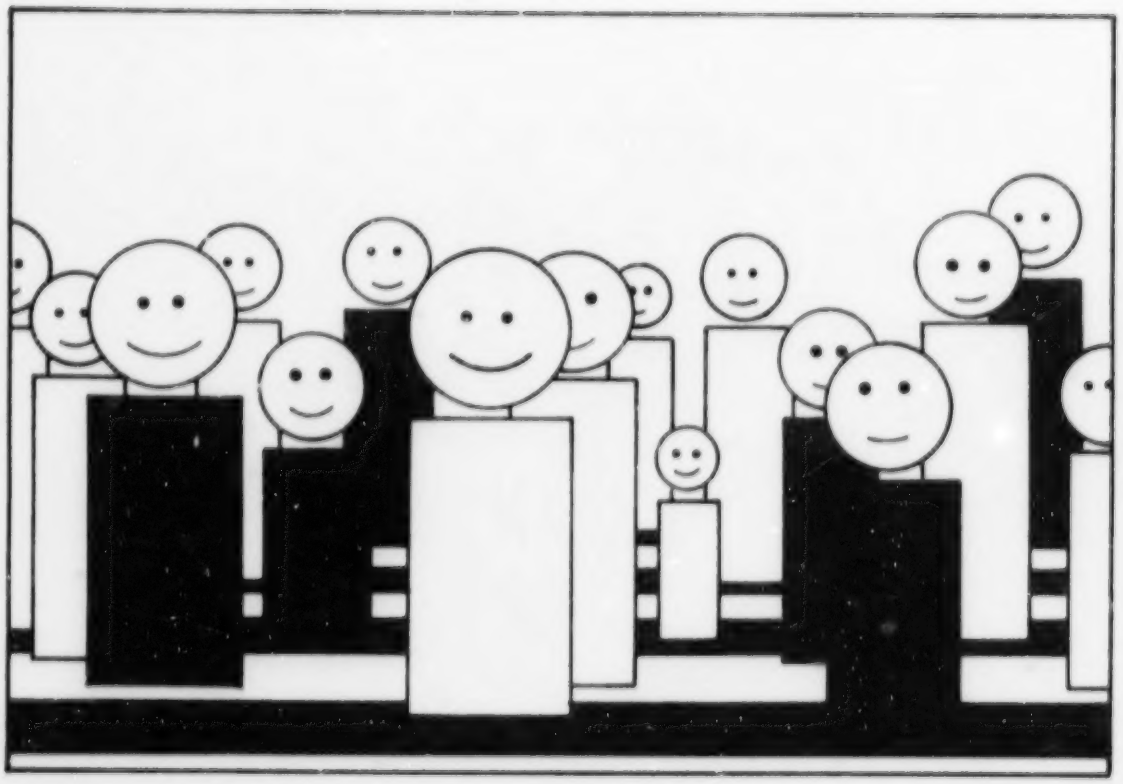
D 6.5 Droite de régression

- *Dans le groupe 2, la droite exprimant la taille des bonshommes en fonction de la longueur de leur ceinture est définie par l'équation:*

  taille = -2 x longueur de ceinture + 16.

Une droite qui permet d'exprimer le mieux possible la liaison entre deux variables.

Population
et échantillon



CONSTATATIONS

C 7.1 Il est relativement facile de déterminer le pourcentage de gens de sexe féminin inscrits à la page 116.

C 7.2 Il est difficile, du moins long, de trouver:

i) le pourcentage de gens de sexe féminin, masculin, impossible à identifier sexuellement, à partir de l'annuaire;

ii) le pourcentage de noms francophones, anglophones, autres et impossibles à déterminer dans l'annuaire en parcourant toutes ses pages.

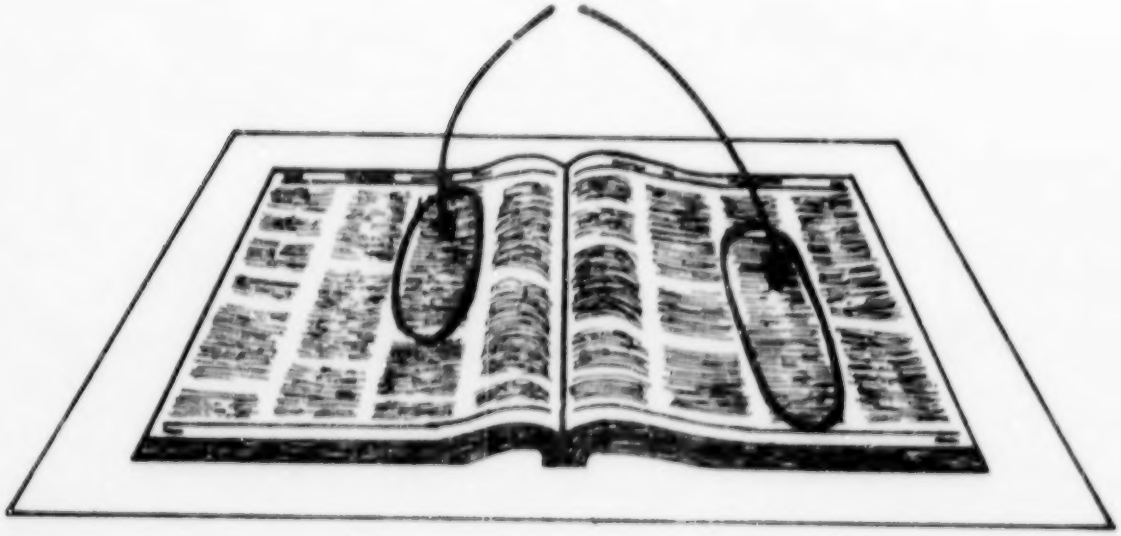
C 7.3 Il est parfois plus facile mais moins précis d'estimer les valeurs cherchées en employant seulement une partie des données.

C 7.4 Il y a plusieurs façons de choisir des données pour estimer:

i) le pourcentage de gens de sexe féminin, masculin, impossible à identifier sexuellement, à partir de l'annuaire;

ii) le pourcentage de noms francophones, anglophones, autres et impossibles à déterminer dans l'annuaire téléphonique.

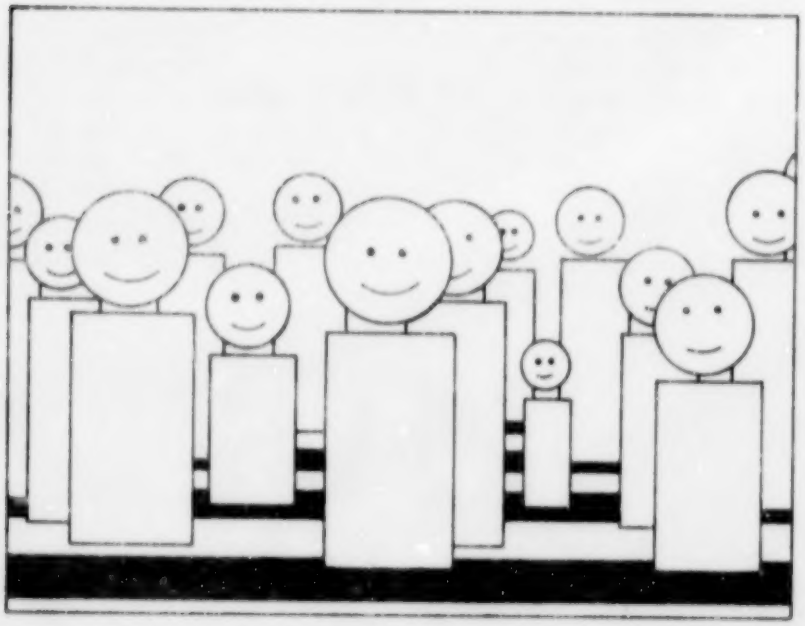
1ère façon



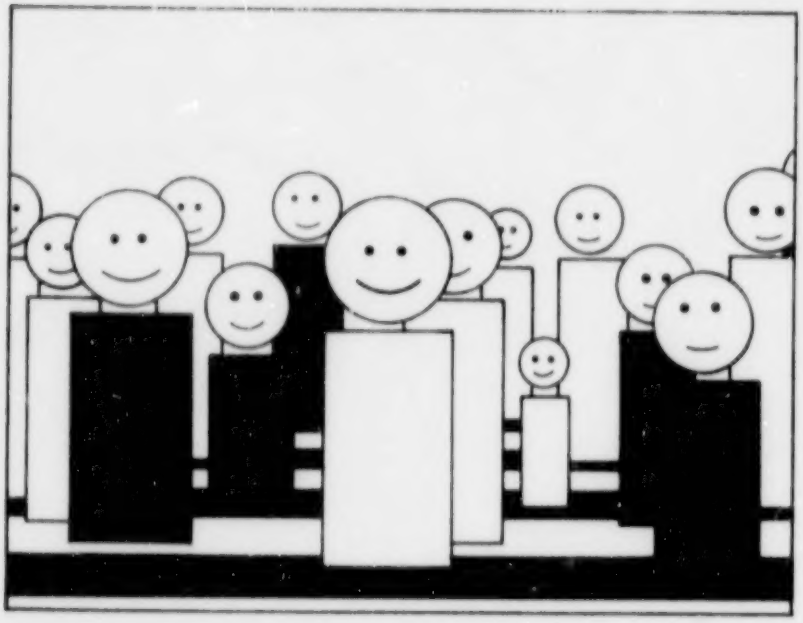
2ième façon

FAITS À RETENIR

F 7.1 On peut travailler parfois facilement, parfois difficilement, avec l'ensemble des données.



F 7.2 On peut estimer la valeur cherchée en n'employant qu'une partie des données.



F 7.3 La précision de notre estimation dépend:

a) du choix des données,
b) de la quantité de données choisies,
c) de la méthode employée.

Il est peut-être préférable d'analyser un plus grand nombre de pages ou de colonnes qu'une ou deux pages ou colonnes successives.

F 7.4 On doit travailler avec toutes les données si on n'accepte aucune erreur d'échantillonnage.

DÉFINITIONS

D 7.1 Unité statistique ou individu

- *Le numéro téléphonique de monsieur X ou de madame Y*
- *Le nom d'une personne*

Ce terme désigne tout élément simple ou composé sur lequel porte l'étude; toute observation qualitative ou quantitative.

D 7.2 Population

- *L'ensemble des personnes à la page 116*
- *L'ensemble des noms dans l'annuaire téléphonique*
- *L'ensemble des gens de la ville où se situe ton institution*

L'ensemble des unités ou individus constituant la collectivité à laquelle on s'intéresse.

D 7.3 Échantillon

- *Le sous-ensemble de noms francophones d'une page de l'annuaire téléphonique.*
- *Le sous-ensemble de gens de sexe féminin de quelques colonnes de l'annuaire téléphonique de la ville où se situe ton institution.*

Groupe d'unités tirées d'une population préalablement définie.

D 7.4 Méthode d'échantillonnage (ou de sondage)

- *Les méthodes employées pendant l'expérience comme*

  i) prendre les numéros de la première et de la deuxième page pour l'estimation;

  ii) prendre les numéros d'une colonne par page.

Les règles visant à construire un échantillon représentatif de la population.

D 7.5 Base d'échantillonnage (de sondage)

- *L'annuaire téléphonique*

Liste, carte ou tout autre ensemble d'informations définissant la population de manière à permettre le tirage des unités de l'échantillon. La base constitue le moyen d'accès réel aux unités de la population.

D 7.6 Taille ou effectif de la population (N)

- *Le nombre de personnes à la page 116*
- *Le nombre de noms dans l'annuaire téléphonique*
- *Le nombre de gens de la ville où se situe ton institution*

Le nombre d'unités statistiques constituant la population.

D 7.7 Taille ou effectif de l'échantillon (n)

- *Le nombre de noms francophones dans l'annuaire téléphonique*
- *Le nombre de gens de sexe féminin dans l'annuaire téléphonique de la ville où se situe ton institution*

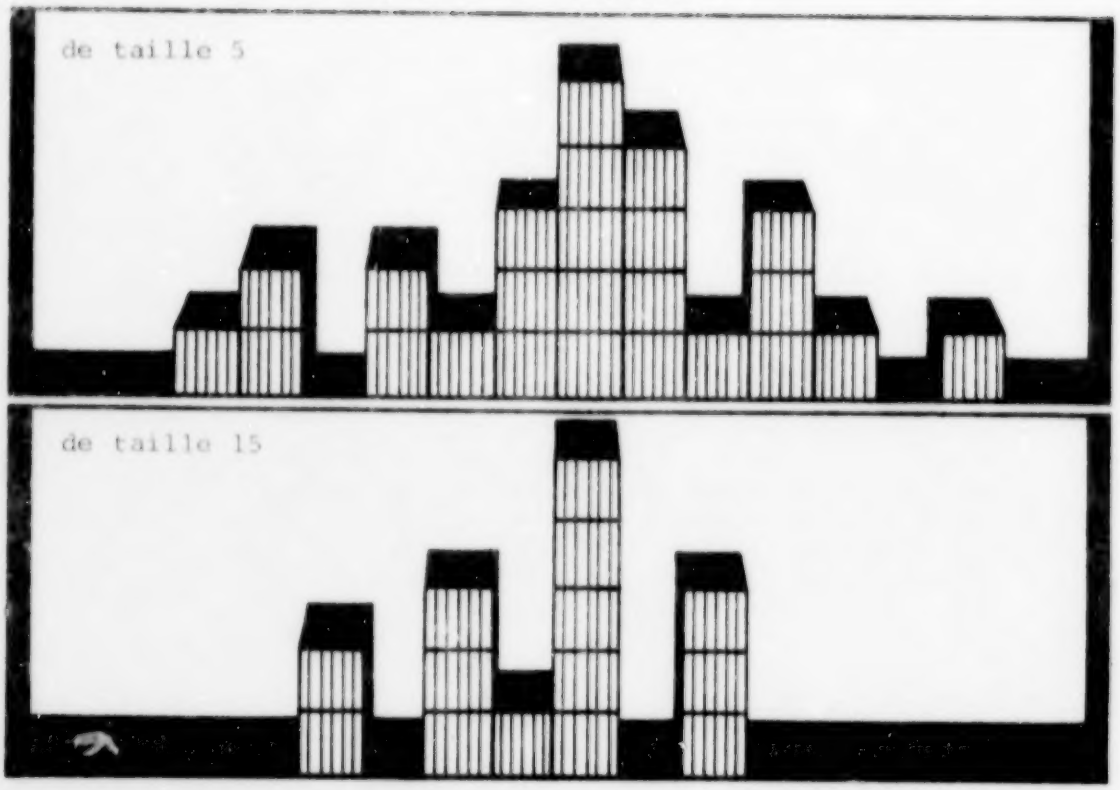
Le nombre d'unités statistiques constituant l'échantillon.

D 7.8 Taux d'échantillonnage (de sondage)

Le rapport de l'effectif de l'échantillon à l'effectif de la population.

Échantillonnage

Distribution des moyennes d'échantillon



CONSTATATIONS

C 8.1 La moyenne d'échantillon varie d'un individu à l'autre.

C 8.2 La moyenne d'échantillon peut être plus ou moins proche, pas nécessairement égale à la moyenne de la population; cela est vrai pour un échantillon de 5 bonshommes et aussi de 15 bonshommes.

C 8.3 La tendance centrale des moyennes d'échantillon de 5 bonshommes est presqu'égale à la tendance centrale des moyennes d'échantillon de 15 bonshommes et presqu'égale à la tendance centrale de la population.

C 8.4 La dispersion des moyennes d'échantillon de 5 bonshommes est plus grande que celle des moyennes d'échantillon de 15 bonshommes.

C 8.5 La dispersion des données de la population est plus grande que celle des moyennes d'échantillon de 5 bonshommes et encore plus grande que celle des moyennes d'échantillon de 15 bonshommes.

FAITS À RETENIR

F 8.1 L'estimation de la moyenne d'une population peut se faire à partir d'un échantillon.

F 8.2 Dans cette estimation plus la taille de l'échantillon est grande, plus la moyenne de l'échantillon a des chances d'être près de la moyenne de la population.

F 8.3 La tendance centrale des moyennes d'échantillon est à peu près égale à la tendance centrale de la population, peu importe le nombre d'éléments de l'échantillon.

F 8.4 La dispersion des moyennes d'échantillon diminue quand la taille de l'échantillon augmente.

F 8.5 La dispersion des moyennes d'échantillon est toujours plus petite que la dispersion des données de la population.

DÉFINITIONS

D 8.1 Distribution des moyennes d'échantillon

- *Tableau des moyennes d'échantillon de 5 bonshommes.*
- *Tableau des moyennes d'échantillon de 15 bonshommes.*

La distribution de toutes les moyennes d'échantillon tiré d'une population.

D 8.2 Tendance centrale des moyennes d'échantillon

- *La tendance centrale des moyennes d'échantillon de 5 bonshommes.*
- *La tendance centrale des moyennes d'échantillon de 15 bonshommes.*

Une valeur représentative de toutes les moyennes d'échantillon de même taille.

D 8.3 Dispersion des moyennes d'échantillon

- *La variation des moyennes d'échantillon (goujons sur le tableau) de taille 5.*
- *La variation des moyennes d'échantillon de taille 15.*

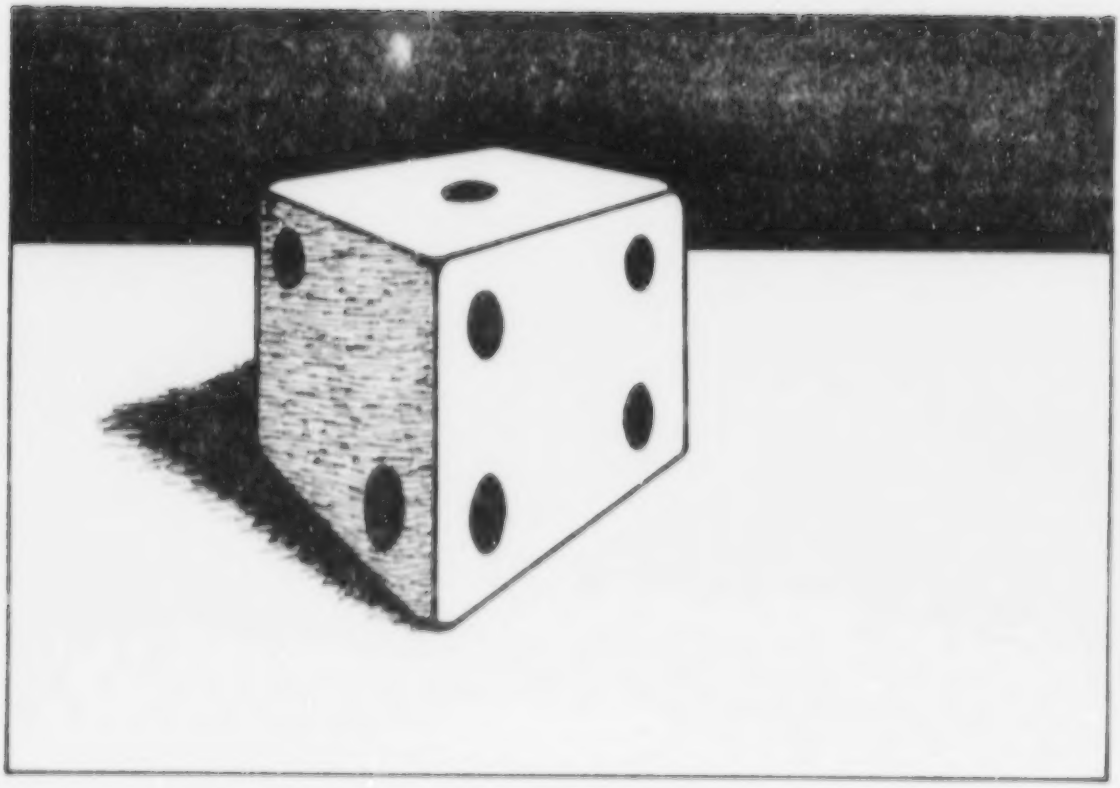
La variation des différentes moyennes d'échantillon de même taille.

D 8.4 Estimation

- *Estimation de la tendance centrale de la population à partir d'un échantillon de taille 5.*
- *Estimation de la tendance centrale de la population à partir d'un échantillon de taille 15.*

Extension à une population de l'information recueillie d'un échantillon.

Probabilité



CONSTATATIONS

C 9.1 Il y a des événements dont la réalisation est

i) certaine

- avoir quelqu'un aux yeux noirs dans le groupe 4;

ii) impossible

- avoir quelqu'un aux cheveux gris dans le groupe 5;

iii) incertaine

- avoir quelqu'un aux cheveux bruns dans le groupe 1.

C 9.2 La chance de réaliser un événement augmente s'il y a plusieurs façons de le réaliser.

- La chance de retirer un bonhomme aux cheveux bruns est plus élevée dans le groupe 9 que dans le groupe 6.

C 9.3 La chance de réaliser un événement diminue si les contraintes augmentent.

- Dans le groupe 2, la chance d'avoir:

i) quelqu'un de taille 22 est 4/6;

ii) quelqu'un de taille 22 et de sexe féminin est 2/6;

iii) quelqu'un de taille 22 et de sexe féminin et aux cheveux gris est 1/6.

C 9.4 Deux ou plusieurs événements lors d'une même épreuve peuvent être réalisés en même temps.

- Avoir quelqu'un de taille 22 et avoir quelqu'un de sexe féminin dans le groupe 2.

- Avoir quelqu'un de taille 22, avoir quelqu'un de sexe féminin, avoir quelqu'un aux cheveux gris, avoir quelqu'un aux bras longs, dans le groupe 2.

C 9.5 Le changement de contexte peut ou ne peut pas influencer la probabilité de réaliser un événement.

- Dans le groupe 2, la probabilité d'avoir quelqu'un aux cheveux gris est 1/6 et la probabilité d'avoir quelqu'un aux cheveux gris sachant qu'il a les bras longs est 1/2.

- Dans le groupe 2, la probabilité d'avoir quelqu'un aux bras longs est 2/6 et la probabilité d'avoir quelqu'un aux bras longs sachant qu'il a les cheveux gris est 1.

- Dans le groupe 8, la probabilité d'avoir quelqu'un aux bras longs est 2/6 et la probabilité d'avoir quelqu'un aux bras longs sachant qu'il a les cheveux bruns est 1/3.

FAITS À RETENIR

F 9.1 La probabilité d'un événement dépend:

1) de l'épreuve subie,
2) de la quantité de résultats favorables à sa réalisation,
3) de la quantité de contraintes impliquées dans sa réalisation,
4) de la quantité de résultats possibles,
5) du contexte possiblement.

F 9.2 La réalisation d'un événement est soit:

1) certaine,
2) impossible,
3) incertaine.

ÉVÉNEMENT: E			
Réalisation	Impossible	Possible	
		Incertaine	Certaine
Probabilité	0	$0 < Pr(E) < 1$	1

F 9.3 La probabilité d'un événement certain est 1.

F 9.4 La probabilité d'un événement impossible à réaliser est 0.

F 9.5 La probabilité d'un événement dont la réalisation est incertaine est un nombre entre 0 et 1.

F 9.6 La probabilité d'un événement peut s'évaluer en faisant le rapport des résultats favorables et des résultats possibles.

F 9.7 Lors d'une épreuve,

i) deux ou plusieurs événements peuvent être réalisés simultanément;

ii) deux ou plusieurs événements peuvent se réaliser sans s'affecter mutuellement, c'est-à-dire d'une façon indépendante.

F 9.8 Lors d'une épreuve,

on peut chercher la réalisation d'un événement en imposant des conditions (ou contexte).

DÉFINITIONS

D 9.1 Épreuve aléatoire (ou expérience aléatoire)

- *Tirer un bonhomme dans le groupe 4.*
- *Tirer un bonhomme dans le groupe 5.*
- *Tirer un bonhomme dans le groupe 6.*

Expérience ou processus dont le(s) résultat(s) est(sont) incertain(s) bien que connu(s) ou connaissable(s).

D 9.2 Ensemble fondamental (S)

- *L'ensemble des bonshommes du groupe 4*
- *L'ensemble des bonshommes du groupe 5*
- *L'ensemble des bonshommes du groupe 2*

Ensemble des résultats possibles de l'expérience aléatoire.

D 9.3 Événement (E_i)

- *Avoir les yeux noirs.*
- *Avoir les cheveux gris.*
- *Avoir la taille 21.*
- *Être de sexe féminin.*
- *Avoir les bras longs.*
- *Avoir une ceinture rouge.*

Action, état ou réalisation désirée dans l'épreuve.

D 9.4 Événement impossible ($\emptyset$)

- *Avoir les cheveux gris si tu prends un bonhomme du groupe 5.*

Événement dont la chance d'être réalisé est nulle; aucun résultat de l'épreuve ne le réalise.

D 9.5 Événement certain (S)

- *Avoir les yeux noirs si tu prends un bonhomme du groupe 4.*
- *Avoir les bras longs si tu prends un bonhomme du groupe 2 qui a les cheveux gris.*

Événement dont la chance d'être réalisé est certaine; tout résultat de l'épreuve se réalise.

D 9.6 Événement composé

- *La taille 22 et le sexe féminin*
- *La taille 22 et le sexe féminin et les cheveux gris*
- *Les cheveux noirs ou les bras courts*

Événement pouvant être décrit comme la réunion ou l'intersection d'événements.

D 9.7 Événements incompatibles

- *Avoir la taille 22 et le sexe féminin et les cheveux gris et les bras longs et la ceinture rouge si tu prends un bonhomme du groupe 2.*

Événements impossibles à réaliser simultanément; aucun résultat de l'épreuve ne les réalise en même temps.

D 9.8 Événements indépendants

- *Avoir les bras longs et avoir les cheveux bruns si tu prends un bonhomme dans le groupe 8.*

Événements dont la réalisation de l'un n'affecte pas la chance de réaliser l'autre.

D 9.9 Probabilité d'un événement ($Pr(E_i)$)

- *Probabilité d'avoir les yeux noirs*
- *Probabilité d'avoir les cheveux gris*

La chance que l'événement se réalise.

D 9.10 Probabilité composée

- *La probabilité de choisir un bonhomme aux cheveux noirs ou aux bras courts dans le groupe 8*
- *La probabilité de choisir un bonhomme de taille 32 et de sexe féminin et aux cheveux gris dans le groupe 7*

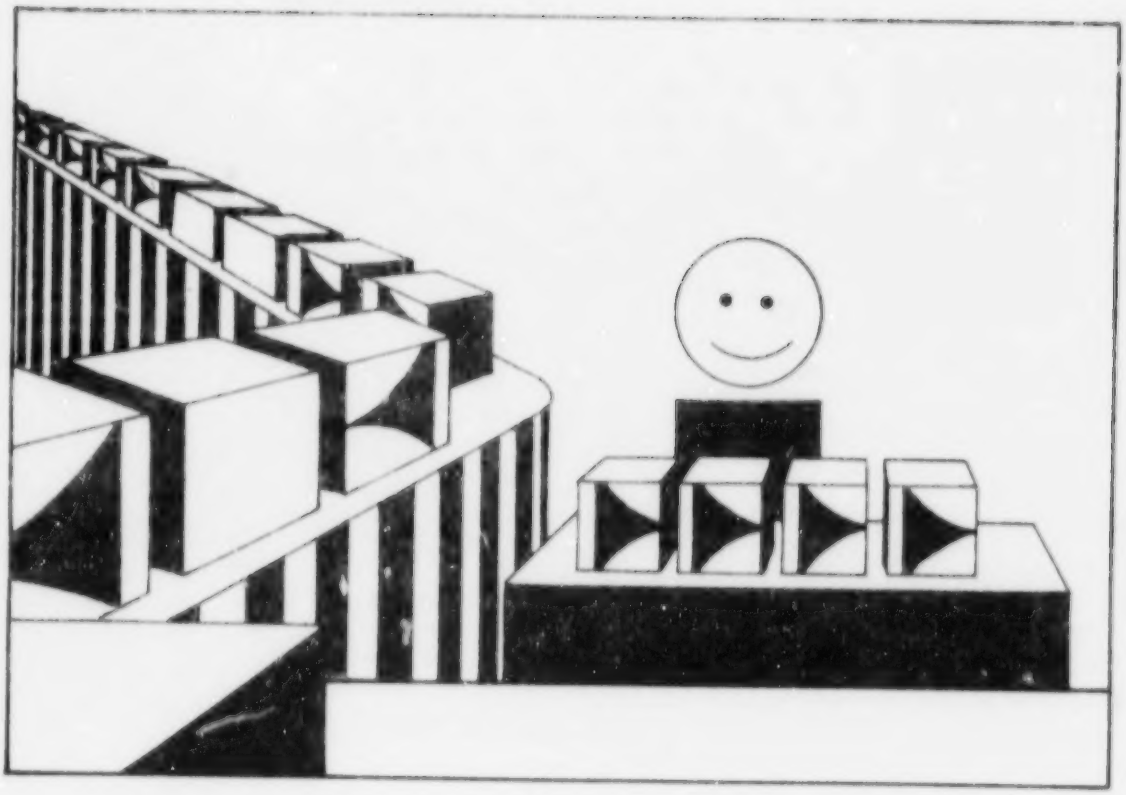
Probabilité de l'événement composé qui implique la réalisation d'un certain nombre de contraintes.

D 9.11 Probabilité conditionnelle ($Pr\ (E_i / E_j)$)

- *La probabilité d'avoir les cheveux gris à condition que les bonshommes du groupe 2 aient les bras longs*
- *La probabilité d'avoir les bras longs à condition que les bonshommes du groupe 2 aient les cheveux gris*

Probabilité de l'événement en tenant compte du fait (ou sous la condition) qu'un autre événement est réalisé.

Erreur d'échantillonnage



CONSTATATIONS

C 10.1 Si je retire un bonhomme dans le groupe 1 et le groupe 7,

a) il est impossible d'avoir un bonhomme de taille
   i) 18 dans le groupe 1,
   ii) 26 dans le groupe 7;

b) il est possible d'avoir un bonhomme de taille 30 dans les 2 groupes mais la chance est plus élevée dans le groupe 7 que dans le groupe 1.

C 10.2 Dans le groupe 1,

a) lorsque je tire deux bonshommes, on a plus de chance d'avoir un échantillon de taille moyenne 26 ou 29 si l'échantillonnage a été fait sans remise qu'avec remise;

b) lorsque je tire trois bonshommes, on a plus de chance d'avoir un échantillon de taille moyenne 28 si l'échantillonnage a été fait sans remise qu'avec remise.

C 10.3 Dans le groupe 7,

a) lorsque je tire deux bonshommes, on a plus de chance d'avoir un échantillon de taille moyenne 18 ou 30 si l'échantillonnage a été fait avec remise que sans remise et c'est le contraire pour un échantillon de taille moyenne 24;

b) lorsque je tire trois bonshommes, on a plus de chance d'avoir un échantillon de taille moyenne 22 ou 26 si l'échantillonnage a été fait sans remise qu'avec remise.

C 10.4 Dans le groupe 1,

a) on a plus de chance d'avoir un échantillon de taille moyenne 28 si je tire un échantillon de 1 ou 2 bonshommes avec remise, de même si j'en tire 3 sans remise;

b) si on accepte une erreur d'une unité dans notre estimation de la moyenne de la population, sans ou avec remise, il est préférable de prendre 2 bonshommes au lieu de 1.

C 10.5 Dans le groupe 7, si on accepte une erreur de deux unités dans notre estimation de la moyenne de la population, il est préférable de prendre, avec ou sans remise, 3 bonshommes au lieu de 2.

C 10.6 Dans le groupe 1, si on tire deux ou trois bonshommes avec remise, on a plus de chance de tomber sur un échantillon dont l'erreur dans l'estimation de la moyenne de la population soit de 2 unités au lieu de 1.

C 10.7 Dans le groupe 7, si on tire deux ou trois bonshommes avec remise, on a plus de chance de tomber sur un échantillon dont l'erreur d'estimation de la moyenne de la population soit de 6 unités au lieu de 3, 2 ou 1.

C 10.8 Dans le groupe 1, si on tire deux bonshommes avec remise, les échantillons différents

(26;30), (28;28), (30;26),

ont la même taille.

C 10.9 Un échantillon composé de 1 ou 2 bonshommes (avec remise), de taille moyenne 26, a plus de chance de provenir du groupe 1 que du groupe 7.

C 10.10 Un échantillon composé de 1, 2 ou 3 bonshommes (avec ou sans remise), de taille moyenne 28, a plus de chance de provenir du groupe 1 que du groupe 7.

C 10.11 Un échantillon composé de 1, 2 ou 3 bonshommes (avec remise), de taille moyenne 30, a plus de chance de provenir du groupe 7 que du groupe 1. Il ne peut provenir ni du groupe 1, ni du groupe 7, s'il est de taille 3 et a été tiré sans remise.

C 10.12 Si l'échantillon composé de 2 bonshommes tirés avec remise a:

i) une taille moyenne de 30 ou 24, nous gagerons qu'il provient du groupe 7 au lieu du groupe 1 et notre chance d'erreur est de 4/13 ou nulle selon le cas;

ii) une taille moyenne de 26, nous gagerons au contraire qu'il provient du groupe 1 au lieu du groupe 7 et la chance de nous tromper est nulle.

FAITS À RETENIR

F 10.1 Énumérer tous les échantillons possibles nous permet d'analyser l'erreur d'échantillonnage.

F 10.2 Le nombre d'échantillons possibles et différents varie selon le mode de tirage (avec ou sans remise) et selon la taille de l'échantillon.

F 10.3 Des échantillons différents peuvent avoir la même taille moyenne.

F 10.4 Lorsqu'on tire un échantillon, la probabilité d'obtenir des bonshommes d'une certaine taille moyenne dépend du groupe, du nombre de bonshommes dans l'échantillon et du mode de tirage.

F 10.5 L'erreur d'échantillonnage varie selon le mode de tirage, la taille de l'échantillon.

F 10.6 Pour estimer la moyenne de la population à partir d'un échantillon, la taille de l'échantillon, l'erreur acceptable (ou la précision désirée) et la chance que cette erreur ne soit pas dépassée sont intimement reliés:

a) si on augmente la taille de l'échantillon, la probabilité de ne pas dépasser l'erreur fixée augmente;

b) si on accepte une plus grande erreur, on a plus de chance de tomber sur un échantillon qui nous permet de ne pas dépasser l'erreur acceptée.

F 10.7 Pour contrôler l'erreur sur un paramètre à estimer à partir d'un échantillon, il nous faut en connaître la distribution d'échantillonnage.

F 10.8 Entre deux éventualités, on choisit naturellement celle qui a le plus de chance d'être réalisée; cependant, il y a toujours une possibilité d'erreur.

F 10.9 On peut toujours évaluer l'erreur possible de notre décision et ainsi la contrôler si on connaît l'ensemble des résultats possibles de chaque éventualité. Dans le cas de la provenance d'un échantillon, on peut faire de même.

DÉFINITIONS

D 10.1 Paramètre

- *La taille moyenne des bonshommes du groupe 1*
- *La taille moyenne des bonshommes du groupe 7*

Mesure faite sur les éléments d'une population.

D 10.2 Statistique

- *La taille moyenne des bonshommes d'un échantillon tiré avec ou sans remise de 3 éléments de la population*
- *La taille moyenne des bonshommes d'un échantillon tiré avec ou sans remise de 4 éléments de la population*

Mesure faite sur les éléments d'un échantillon.

D 10.3 Estimation

- *L'évaluation de la taille moyenne du groupe à partir d'un échantillon de 3 ou 4 bonshommes du groupe 3 tirés avec ou sans remise*

Recherche de la valeur d'un ou plusieurs paramètres d'une population à partir des observations d'un échantillon.

D 10.4 Estimation ponctuelle

- *L'évaluation de la taille moyenne du groupe numéro 1 à partir de la taille moyenne d'un échantillon de 3 ou 4 bonshommes tirés avec ou sans remise, du groupe 1*
- *L'évaluation de la taille moyenne du groupe numéro 7 à partir de la taille moyenne d'un échantillon de 3 ou 4 bonshommes tirés avec ou sans remise, du groupe 7*

Recherche d'une valeur unique, considérée comme optimale, à partir des observations d'un échantillon pour l'estimation d'un paramètre d'une population.

D 10.5 Précision ou erreur d'estimation

- *La valeur: une, deux etc. unités font s'éloigne la taille moyenne de l'échantillon de 2 ou 3 bonshommes tirés avec ou sans remise dans l'estimation de la taille moyenne du groupe 1;*
- *La valeur: une, deux etc. unités dont s'éloigne la taille moyenne de l'échantillon de 2 ou 3 bonshommes tirés avec ou sans remise dans l'estimation de la taille moyenne du groupe 2*

L'erreur maximale qu'on peut faire dans l'estimation d'un paramètre inconnu.

D 10.6 Estimation d'intervalle de confiance

- *Intervalle obtenu en prenant la taille moyenne d'un échantillon de 2 ou 3 bonshommes tirés avec ou sans remise du groupe 1 ou 2 et en lui ajoutant et soustrayant une, deux etc. unités*

Recherche d'un intervalle, à partir des observations d'un échantillon, pour l'estimation d'un paramètre d'une population dont on fixe la probabilité de l'erreur d'estimation.

D 10.7 Niveau de confiance de l'estimation

- *la probabilité 1/3 que la taille moyenne de l'échantillon de deux bonshommes tirés avec remise dans le groupe 1 ne s'éloigne pas plus d'une unité de la taille moyenne des bonshommes du groupe 2*
- *La probabilité 48/64 que la taille moyenne de l'échantillon de trois bonshommes tirés avec remise dans le groupe 2 ne s'éloigne pas plus de deux unités de la taille moyenne des bonshommes du groupe 2*

La probabilité de ne pas dépasser l'erreur d'estimation.

D 10.8 Estimateur

- *La taille moyenne d'un échantillon pour estimer la taille moyenne des bonshommes du groupe 1 ou du groupe 2*

Formule ou méthode de calcul utilisée pour effectuer une estimation ponctuelle à partir des valeurs observées sur l'échantillon.

D 10.9 Décision statistique

- *Miser que l'échantillon de 2 bonshommes tirés avec remise, de taille moyenne 30, provient du groupe 2*
- *Miser que l'échantillon de 2 bonshommes tirés avec remise, de taille moyenne 50, provient du groupe 1*

Décision sur la nature d'une population basée sur l'information obtenue dans l'échantillon.

D 10.10 Critère de décision statistique

- *La probabilité que l'échantillon de deux bonshommes tirés avec remise, de taille moyenne 50, provienne du groupe 1 est 1 tandis que la probabilité de provenir du groupe 2 est nulle*
- *La probabilité que l'échantillon de deux bonshommes tirés avec remise, de taille moyenne 30, provienne du groupe 2 est 0,12 tandis que la probabilité de provenir du groupe 1 est 0,72*

La probabilité de l'information obtenue d'un échantillon par rapport aux différentes hypothèses émises sur la nature d'une population.

D 10.11 Erreur de décision statistique

- *Miser sur le groupe 2 quand c'est le groupe 1 qui est le bon et vice versa*

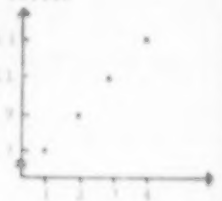
Erreur commise dans la décision prise sur la nature d'une population à partir de l'information obtenue d'un échantillon.

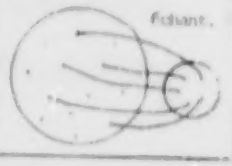
Coffre d'outils

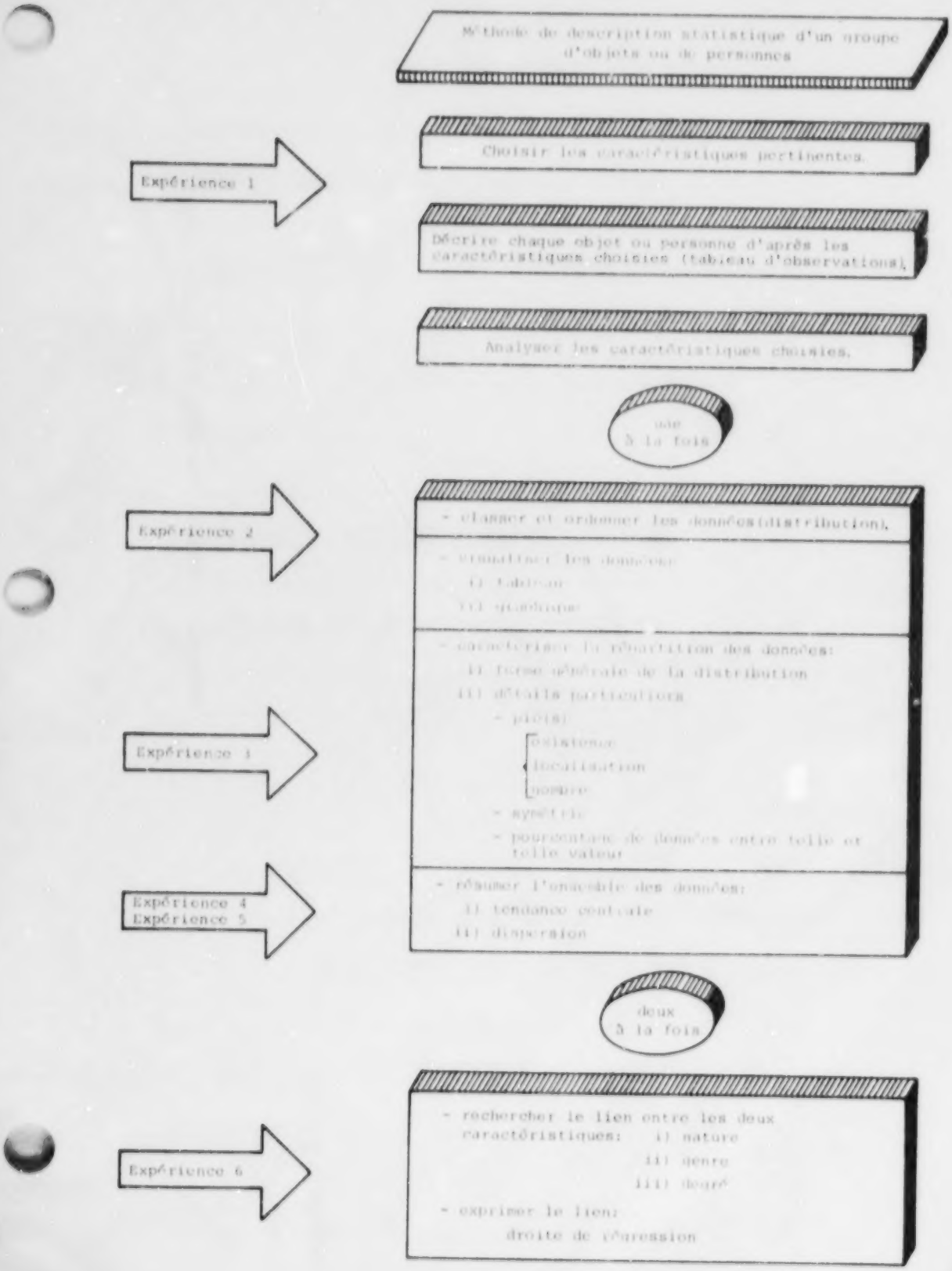
Coffre d'outils

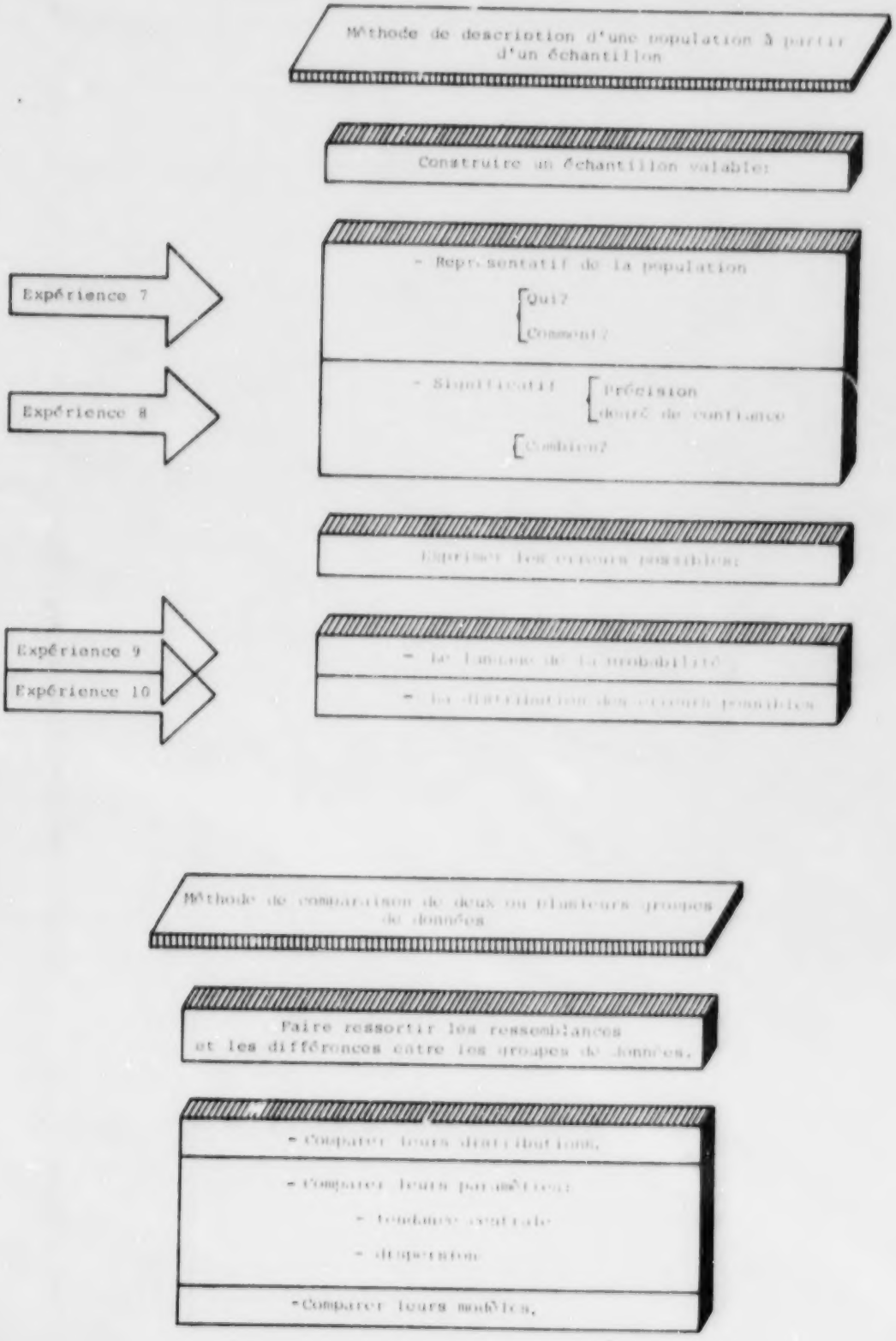
(pour apprenti(e))

SORTE	NOM	UTILISATION	MODE D'EMPLOI	EXEMPLE
Observation	Tableau d'observations	Décrire un groupe par le biais de ses éléments.	- Choisir des caractéristiques pertinentes. - Décrire les individus selon les différentes catégories de chaque caractéristique.	(cm) taille nom / sexe 1 M 162 2 M 167 3 F 150 4 F 160
Classification	Tableau d'effectifs	Décrire un groupe suivant une caractéristique discrétisée.	- Construire un tableau où l'on écrit les différentes catégories (ordonnées si pertinent) de la caractéristique. - Indiquer le nombre d'observations pour chaque catégorie.	CLASSE / EFFECTIF F / 10 M / 15 Total / 25
Visualisation	Représentation graphique		- Écrire les catégories sur un axe horizontal et les effectifs sur un axe vertical. - Représenter les effectifs en utilisant des rectangles ou des points.	Effectif 15 10 5 0 F M Sexe
Identification	Forme de distribution	Reconnaître et comparer des distributions.	- Faire la réunion. - Identifier les pics. - Reconnaître la symétrie. - Déterminer le pourcentage de données entre différentes valeurs de la caractéristique.	Effectif 5 Taille Effectif 10 5 0 Taille
Mesure de tendance centrale	Moyenne arithmétique	Mesurer la tendance centrale.	- Additionner toutes les données. - Diviser cette somme par le nombre de données.	Soit 20; 24; 18; 12; 18 Moyenne $\frac{20+24+18+12+18}{5}$ $\frac{112}{5} = 22,4$
	Mode		- Identifier la donnée la plus fréquente.	Mode = 18
	Médiane		- Ordonner les données. - Trouver la valeur au centre.	De croissant les valeurs. 18; 18; 20; 24; 32 Médiane = 20

Mesure de dispersion	Étendue	Mesurer la dispersion.	Trouver la différence entre la plus grande et la plus petite donnée.	Étendue = 32-18 = 14
	Écart-Moyen		Calculer l'écart entre chaque donnée et la moyenne du groupe. Faire la moyenne de ces écarts.	Écart moyen = (|20-22,4| + |24-22,4| + |18-22,4| + |26-22,4| + |18-22,4|)/5 = 22,4/5 = 4,48
Corrélation	Nuage de points	Mesurer le lien entre deux caractéristiques	- Écrire les valeurs d'une caractéristique sur l'axe horizontal et les valeurs de l'autre caractéristique sur l'axe vertical. - Faire correspondre un point à chaque paire de valeurs des caractéristiques.	Taille  Nbre de point.
	Covariance		- Calculer la différence entre chaque donnée et la moyenne du groupe pour chaque caractéristique. - Calculer la moyenne des produits des différences.	Soient les paires de valeurs (1;2), (2;3), (3;5) et (4;10) Cov = [(1-2,5)(2-5) + (2-2,5)(3-5)+ (3-2,5)(5-5)+ (4-2,5)(10-5)]/4 = 2,5
	Droite de régression	Exprimer le lien entre deux caractéristiques	- Calculer la pente et l'ordonnée à l'origine à partir de deux points lorsque tous les points sont alignés. - Écrire l'équation de la droite.	$a = \frac{y_2 - y_1}{x_2 - x_1}$ x $b = y_1 - 2 \times x_1$ ou $b = 5 - 2 \times 2$ $y = 2x + 1$

Estimation et vérification d'hypothèse	**Échantillon**	- Estimer une population. - Vérifier une hypothèse concernant une population.	- Déterminer la taille de l'échantillon. - Choisir des éléments représentatifs de la population.	Population  échant.
	Distribution d'échantillonnage des moyennes	Montrer les liens entre la moyenne de la population et la moyenne d'un échantillon. Ces liens sont: 1° La tendance centrale de la population égale la tendance centrale des moyennes d'échantillon peu importe la taille des échantillons. 2° La dispersion des moyennes d'échantillon est plus petite que la dispersion de la population; de plus, elle diminue lorsque la taille des échantillons augmente	- Sortir de la population tous les échantillons d'une taille donnée. - Calculer la moyenne de chaque échantillon.	Une population est composée des valeurs 26; 28; 30. On tire avec remise un échantillon de taille deux. Échantillon possible | moyenne 26; 26 | 26 26; 28 | 27 26; 30 | 28 28; 26 | 27 28; 28 | 28 28; 30 | 29 30; 26 | 28 30; 28 | 29 30; 30 | 30 moyenne de la population = 28 moyenne des moyennes d'échantillon = 28 écart moyen de la population $\simeq 1,33$ écart moyen des moyennes d'échantillon $= \frac{8}{9} \simeq 0,89$
Mesure du hasard	**Probabilité**	Évaluer les chances de réaliser un événement.	- Trouver les cas possibles. - Identifier les cas favorables à la réalisation de l'événement. - Additionner les chances de réaliser chaque cas favorable.	On lance un dé une fois. On cherche la chance d'avoir un nombre pair. Les cas possibles sont: 1; 2; 3; 4; 5; 6. Les cas favorables sont 2; 4; 6. La chance d'avoir un nombre pair est: $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$
Analyse d'erreur d'échantillonnage	**Distribution d'échantillonnage des moyennes** **Probabilité**	Évaluer et contrôler l'erreur d'échantillonnage	- Sortir de la population tous les échantillons possibles d'une taille donnée. - Calculer la moyenne de chaque échantillon. - Évaluer l'erreur d'échantillonnage.	Une population est composée des valeurs 25; 28; 30. La chance que la moyenne d'un échantillon de taille deux, tiré avec remise, soit à plus d'une unité de la moyenne de la population égale 2/9.


Méthode de description statistique d'un groupe d'objets ou de personnes
Choisir les caractéristiques pertinentes.
Décrire chaque objet ou personne d'après les caractéristiques choisies (tableau d'observations).
Analyser les caractéristiques choisies.
une à la fois
- classer et ordonner les données(distribution).
- résumer les données:
i) tableau
ii) graphique
- caractériser la répartition des données:
i) forme générale de la distribution
ii) détails particuliers
- pics:
existence
localisation
nombre
- symétrie
- pourcentage de données entre telle et telle valeur
- résumer l'ensemble des données:
i) tendance centrale
ii) dispersion
deux à la fois
- rechercher le lien entre les deux caractéristiques: i) nature
ii) genre
iii) degré
- exprimer le lien:
droite de régression
Expérience 1
Expérience 2
Expérience 3
Expérience 4
Expérience 5
Expérience 6


Méthode de description d'une population à partir d'un échantillon
Construire un échantillon valable:
- Représentatif de la population
Qui?
Comment?
- Significatif
Précision
degré de confiance
Combien?
Expérience 7
Expérience 8
Expérience 9
Expérience 10

Méthode de comparaison de deux ou plusieurs groupes de données
Faire ressortir les ressemblances et les différences entre les groupes de données.
- Comparer leurs distributions.
- Comparer leurs paramètres:
- tendance centrale
- dispersion
- Comparer leurs modèles.

Laboratoire d'applications

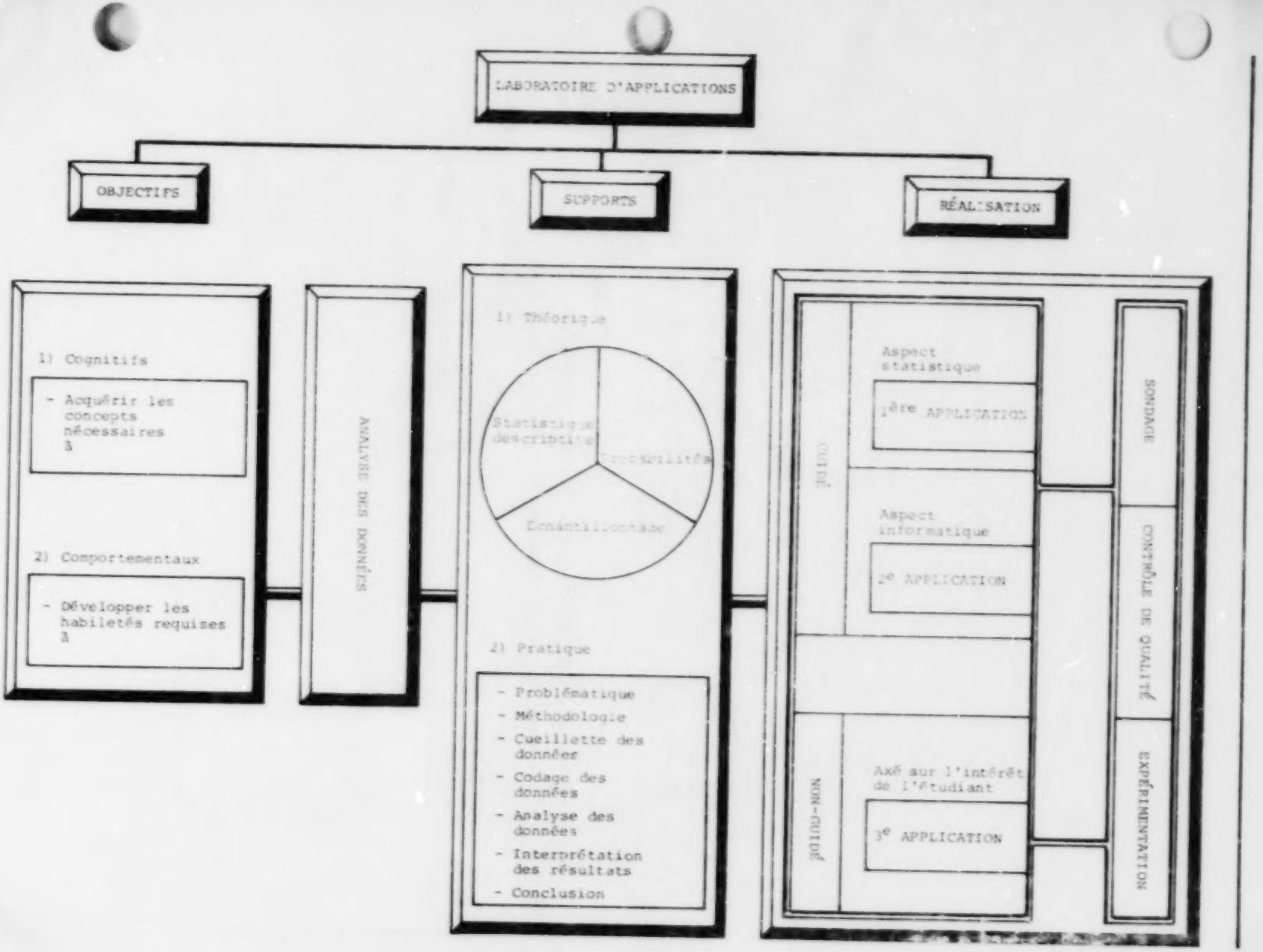

LABORATOIRE D'APPLICATIONS
OBJECTIFS
SUPPORTS
RÉALISATION
1) Cognitifs
- Acquérir les concepts nécessaires à
2) Comportementaux
- Développer les habiletés requises à
ANALYSE DES DONNÉES
1) Théorique
Statistique descriptive
Probabilités
Échantillonnage
2) Pratique
- Problématique
- Méthodologie
- Cueillette des données
- Codage des données
- Analyse des données
- Interprétation des résultats
- Conclusion
GUIDÉ
Aspect statistique
1ère APPLICATION
Aspect informatique
2e APPLICATION
NON-GUIDÉ
Axé sur l'intérêt de l'étudiant
3e APPLICATION
SONDAGE
CONTRÔLE DE QUALITÉ
EXPÉRIMENTATION

Table des matières

1. STATISTIQUE DESCRIPTIVE

- APPLICATION I

  Le rapport du commissaire industriel de la ville de Bécart.

- APPLICATION II

  Les activités étudiantes du groupe au cours des dernières vacances estivales.

2. THÉORIE D'ÉCHANTILLONNAGE

- APPLICATION III

  Étude des plaintes formulées par les clients de la coopérative d'alimentation sur la qualité des pommes de terre.

- APPLICATION IV

  Étude du revenu des étudiants(es) en statistique au cours des dernières vacances estivales.

3. TRAVAUX PRATIQUES

T.1 STRUCTURE TEMPORELLE

Étude de la consommation électrique

T.2 SYNTHÈSE

Étude d'une échelle de salaire

T.3 CORRÉLATION

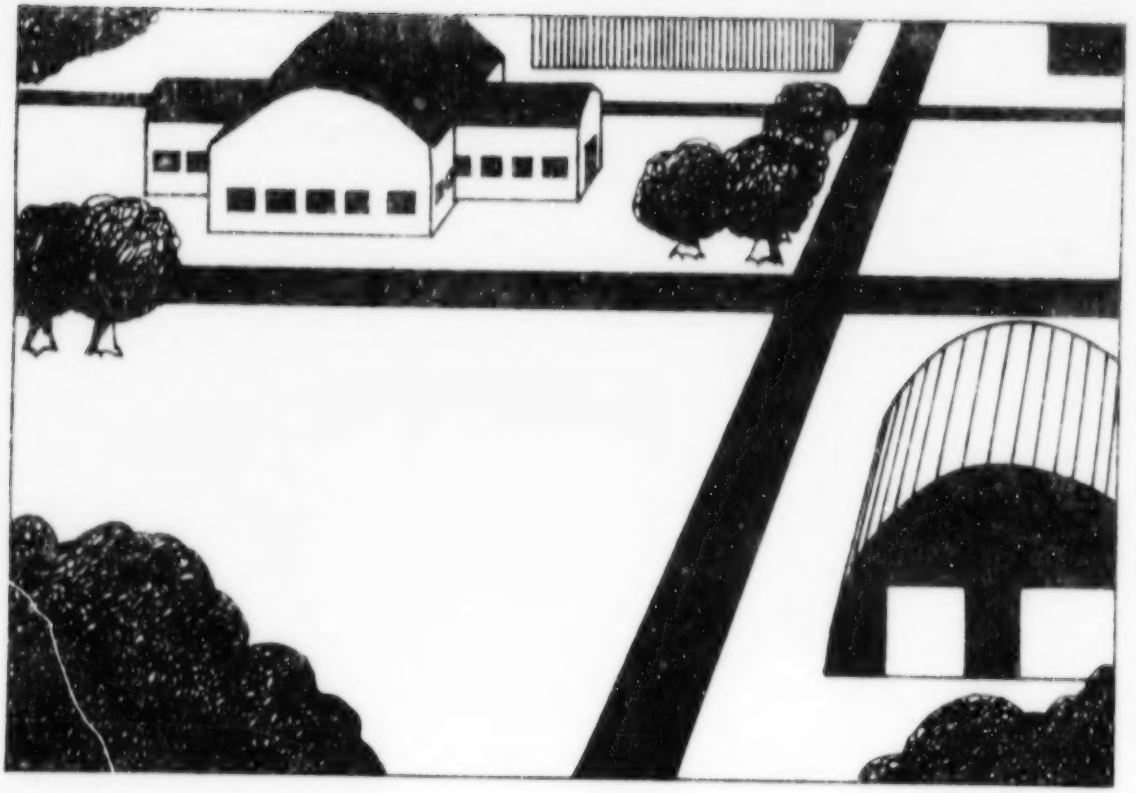
1. Les accidents de la circulation au Québec
2. Les échelles salariales du personnel enseignant

T.4 PROBABILITÉ

1. Le jeu de la Quotidienne
2. Le jeu de la Mini-loto
3. La règle des parties

1. Statistique descriptive

Le rapport du commissaire industriel de la ville de Bécart



APPLICATION 1

NOM(S) : _____

_____

_____

OBJECTIF GÉNÉRAL

Acquérir les concepts statistiques et développer les habiletés nécessaires à l'analyse de données.

MATÉRIEL

Aucun.

CONTEXTE DE L'APPLICATION

Nous voulons étudier la structure salariale et la structure d'âge des employés(es) de quatre compagnies de la ville de Bécart. Cette étude fait appel à la statistique descriptive.

Le commissaire industriel de la ville de Bécart, désireux de promouvoir les intérêts de sa ville et d'y implanter de nouvelles industries, prépare à cet effet un dossier pour le ministère de l'Industrie et du Commerce.

En autres, il veut inclure dans son dossier des données objectives concernant la structure salariale des employés(es) des quatre plus importantes compagnies de sa petite ville, la compagnie Riki inc., BBGR inc., Les Mines Leclerc inc., La Boulangerie Béland. En effet, le salaire payé aux employés(es) des grosses compagnies détermine souvent le niveau de salaire exigé et obtenu par les employés(es) des autres secteurs.

De plus la connaissance de la structure d'âge des employés(es) permanents(es) peut aider à estimer le nombre d'emplois disponibles pour les prochaines années dans ces compagnies.

D'après le contexte historique et industriel de Bécart, les employés(es) les plus vieux(vieilles) sont habituellement les moins spécialisés(es) et les moins bien rémunérés(es). Il s'agit de vérifier ces faits.

En projettant un taux d'inflation de 10%, quel serait le salaire horaire moyen gagné par les employés(es) de ces compagnies pour les 5 prochaines années s'ils obtiennent la pleine indexation? La structure de leur salaire serait-elle modifiée? L'écart entre les faibles salariés et les hauts salariés diminuerait-il ou augmenterait-il?

Ces interrogations suscitent l'intérêt de tout futur investisseur; le commissaire industriel se propose de faciliter la tâche de l'investisseur et il te confie le soin de répondre à ces interrogations.

Pour t'aider, le commissaire industriel te remet la liste des employés(es) de chaque compagnie ainsi que leur salaire horaire et leur âge (tableaux 1, 2, 3 et 4).

A partir de cette information, tu peux trouver la structure salariale, la structure d'âge, étudier les relations entre l'âge et le salaire horaire des employés(es) et extrapoler les salaires pour les cinq prochaines années.

1. L'étude de la structure salariale des employés(es)

Trouver la distribution du salaire horaire des employés(es) (classer et ordonner) et représenter cette distribution (tableau ou graphique).

i) La compagnie Riki inc.

ii) La compagnie BBGR inc.

iii) Les Mines Leclerc inc.

iv) La Boulangerie Béland

Comparer les distributions du salaire horaire des employé(es) des quatre compagnies. (Les quatre groupes d'employés(es) sont-ils payés de la même façon?)

Synthétiser chaque distribution par des paramètres qui renseignent sur la tendance générale et la variation du salaire horaire de chaque groupe d'employés(es).

Paramètre	Nom des compagnies			
	Riki inc.	BBCR inc.	Les Mines Leclerc inc.	La Boulangerie Béland
Tendance générale				
Variation				

Interpréter et comparer les résultats. (Quel groupe gagne le plus, le moins? Les salaires varient-ils de la même façon d'un groupe à l'autre? etc.)

Chercher, pour chaque compagnie, un modèle qui exprime relativement bien la structure salariale des employés(es). Expliquer. (Chercher une forme caractéristique pour la distribution.)

i) La compagnie Riki inc.

ii) La compagnie BBGR inc.

iii) Les Mines Leclerc inc.

iv) La Boulangerie Béland

Mentionner les aspects les plus intéressants de la structure salariale des employés(es).

2. L'étude de la structure d'âge des employés(es)

Étudier la structure d'âge des employé(es) comme la structure salariale c'est-à-dire chercher d'abord la distribution d'âge des employé(es), les paramètres qui la résument et un modèle si possible qui la représente, pour finalement en dégager les aspects les plus intéressants.

3. **L'étude des relations entre le salaire horaire et l'âge des employés(es)**

Comparer, pour chaque compagnie, la structure salariale et la structure d'âge des employé(es).

i) La compagnie Riki inc.

ii) La compagnie BBGR inc.

iii) Les Mines Leclerc inc.

iv) La Boulangerie Béland

Trouver la distribution conjointe du salaire horaire et de l'âge des employé(es) (classer et ordonner) et représenter cette distribution (tableau à double entrée ou nuage de points).

i) La compagnie Riki inc.

ii) La compagnie BBGR inc.

iii) Les Mines Leclerc inc.

iv) La Boulangerie Béland

Comparer les distributions conjointes du salaire horaire et de l'âge des employé(e)s des quatre compagnies.

(Les quatre groupes d'employé(e)s sont-ils payés de la même façon selon leur âge.)

Trouver, s'il existe, le lien entre le salaire horaire et l'âge des employés(es). (Calculer le coefficient de corrélation linéaire entre le salaire horaire et l'âge des employés(es).)

	Nom des compagnies			
	Riki inc.	BBGR inc.	Les Mines Leclerc inc.	La Boulangerie Béland
Coefficient de corrélation linéaire				

Interpréter les résultats obtenus.

Exprimer le lien, s'il existe, entre le salaire horaire et l'âge des employé(es). (Trouver la droite de régression linéaire.)

i) La compagnie Riki inc.

ii) La compagnie BBCR inc.

iii) Les Mines Leclerc inc.

iv) La boulangerie Béland

Mentionner les relations les plus intéressantes entre le salaire horaire et l'âge des employés(es) et les particularités de chaque compagnie.

4. L'étude des prévisions salariales pour les cinq prochaines années

Extrapoler le salaire horaire moyen des employéres, pour les 5 prochaines années, sur la base de la pleine indexation du salaire c'est-à-dire de l'entière compensation de la perte du pouvoir d'achat dû au taux d'inflation annuel évalué à 10%.

Salaire horaire moyen ($)	Année				
	1	2	3	4	5
Riki inc.					
BBGR inc.					
Les Mines Leclerc inc.					
La Boulangerie Béland					

Commenter les résultats.

Calculer la variation du salaire horaire des employé(e)s de chaque compagnie, pour les cinq prochaines années, sur la base de la pleine indexation.

Variation du salaire horaire ($)	Année				
	1	2	3	4	5
Riki inc.					
BBGR inc.					
Les Mines Leclerc inc.					
La Boulangerie Béland					

Commenter les résultats.

Dégager les effets importants de ces prévisions sur la structure salariale des employé(es). (La structure salariale s'est-elle modifiée? L'écart entre les salariés(es) a-t-il changé? Les différences entre les structures salariales des compagnies ont-elles diminué ou augmenté? etc.)

Donner les résultats de l'étude qui répondent aux interrogations de l'investigateur.

Tableau 1 Le salaire horaire et l'âge des employés(es) de la compagnie Riki inc.

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
1	7,00	40	31	7,00	49	61	9,00	29	91	8,00	38	121	9,00	27
2	10,00	18	32	7,00	40	62	8,00	37	92	7,00	46	122	7,00	45
3	8,00	35	33	6,00	55	63	7,00	48	93	9,00	25	123	9,00	28
4	8,00	35	34	8,00	37	64	10,00	19	94	7,00	44	124	7,00	45
5	8,00	30	35	9,00	25	65	8,00	36	95	7,00	45	125	7,00	48
6	6,00	55	36	8,00	25	66	7,00	43	96	7,00	46	126	10,00	19
7	9,00	24	37	7,00	40	67	8,00	34	97	8,00	39	127	6,00	55
8	7,00	44	38	8,00	35	68	6,00	57	98	6,50	54	128	8,00	35
9	7,00	44	39	9,00	24	69	8,00	35	99	5,00	65	129	8,00	35
10	8,00	32	40	7,00	41	70	7,00	44	100	8,00	35	130	9,00	25
11	9,00	25	41	9,00	20	71	8,00	35	101	7,00	45	131	7,00	45
12	6,00	56	42	7,00	42	72	9,00	25	102	6,00	53	132	7,00	45
13	7,00	48	43	9,00	21	73	6,00	56	103	7,00	48	133	8,00	36
14	8,00	36	44	7,00	42	74	7,00	45	104	6,00	58	134	6,00	56
15	7,00	46	45	7,00	43	75	8,00	31	105	8,00	35	135	7,00	41
16	9,00	26	46	7,00	44	76	7,00	46	106	6,00	59	136	8,00	37
17	8,00	37	47	7,00	44	77	9,00	21	107	7,00	49	137	8,00	37
18	5,00	64	48	5,00	61	78	8,00	32	108	9,00	25	138	8,00	36
19	7,00	48	49	7,00	45	79	6,00	52	109	8,00	35	139	7,00	40
20	7,00	42	50	8,00	35	80	8,00	38	110	6,00	55	140	8,00	35
21	7,00	43	51	7,00	45	81	7,00	45	111	7,00	44	141	5,00	65
22	8,00	39	52	7,00	46	82	8,00	35	112	6,00	54	142	7,00	45
23	8,00	35	53	9,00	25	83	6,00	55	113	7,00	45	143	6,00	55
24	6,00	54	54	8,00	39	84	7,00	45	114	9,00	26	144	7,00	47
25	7,00	45	55	7,00	45	85	9,00	25	115	7,00	45	145	9,00	26
26	9,00	25	56	9,00	25	86	7,00	47	116	7,00	45	146	8,00	35
27	7,00	45	57	7,00	45	87	7,00	45	117	9,00	24	147	8,00	32
28	8,00	34	58	7,00	46	88	8,00	38	118	8,00	35	148	6,00	54
29	7,00	46	59	7,00	47	89	8,00	38	119	8,00	35	149	7,00	46
30	7,00	47	60	8,00	35	90	6,00	55	120	7,00	43	150	8,00	31

Tableau 1 (suite) Le salaire horaire et l'âge des employés(es) de la compagnie Riki inc.

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
151	9,00	25	161	8,00	45	171	7,00	44	181	8,00	35	191	7,00	47
152	7,00	45	162	6,00	55	172	8,00	31	182	6,00	55	192	9,00	25
153	8,00	30	163	8,00	45	173	8,00	35	183	7,00	46	193	8,00	37
154	6,00	56	164	8,00	34	174	6,00	57	184	8,00	35	194	6,00	55
155	8,00	39	165	9,00	25	175	8,00	38	185	8,00	52	195	7,00	49
156	9,00	25	166	7,00	45	176	7,00	37	186	9,00	20	196	8,00	35
157	7,00	45	167	6,00	57	177	9,00	25	187	8,00	41	197	6,00	55
158	8,00	35	168	8,00	35	178	8,60	38	188	8,00	34	198	8,00	36
159	6,00	55	169	10,00	16	179	8,00	37	189	6,00	56	199	8,00	35
160	8,00	36	170	8,00	41	180	8,00	34	190	8,00	35	200	7,00	45

Tableau 2 Le salaire horaire et l'âge des employés(es) de BBGR inc.

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
1	7,80	60	31	6,60	55	61	7,80	25	91	6,00	50
2	6,60	30	32	7,80	60	62	9,00	40	92	6,60	30
3	6,00	18	33	9,00	60	63	6,00	50	93	7,20	20
4	6,00	19	34	7,20	50	64	6,60	30	94	7,80	40
5	9,00	25	35	6,60	55	65	7,80	40	95	8,40	35
6	8,40	20	36	9,00	45	66	7,20	45	96	6,00	35
7	7,80	25	37	8,40	50	67	6,00	20	97	6,60	55
8	7,20	20	38	9,00	45	68	9,00	47	98	9,00	25
9	6,00	65	39	8,40	35	69	9,00	46	99	6,60	55
10	8,40	35	40	7,20	60	70	7,80	60	100	6,00	64
11	9,00	40	41	7,80	45	71	8,40	35	101	8,40	65
12	7,80	40	42	7,80	60	72	7,20	65	102	6,00	20
13	7,20	25	43	7,20	35	73	7,80	25	103	8,40	60
14	6,60	30	44	6,00	35	74	6,60	55	104	8,40	61
15	7,80	45	45	8,40	51	75	7,80	46	105	6,60	55
16	6,00	65	46	8,40	65	76	6,00	63	106	9,00	40
17	9,00	45	47	7,80	25	77	9,00	42	107	8,40	25
18	6,60	31	48	6,60	55	78	7,20	50	108	7,80	60
19	7,20	30	49	6,00	20	79	7,80	60	109	7,80	25
20	6,00	64	50	6,60	30	80	9,00	45	110	6,60	30
21	9,00	40	51	8,40	64	81	7,20	55	111	7,20	65
22	8,40	20	52	9,00	45	82	6,00	35	112	6,00	65
23	7,20	35	53	9,00	40	83	6,60	30	113	7,20	40
24	6,00	20	54	6,60	57	84	6,60	55	114	8,40	20
25	7,20	50	55	6,60	30	85	8,40	50	115	7,80	25
26	7,80	25	56	6,00	35	86	9,00	41	116	7,20	35
27	7,20	30	57	7,20	35	87	8,40	20	117	6,60	30
28	8,40	65	58	9,00	45	88	7,20	50	118	9,00	40
29	7,80	60	59	6,00	50	89	9,00	60	119	8,40	50
30	7,20	55	60	6,00	51	90	8,40	25	120	6,60	55

Tableau 3 Le salaire horaire et l'âge des employés(es) des Mines Leclerc inc.

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
1	9,50	60	36	9,50	55	71	8,50	50	106	8,50	50
2	7,00	25	37	9,50	60	72	7,00	35	107	7,00	30
3	11,00	65	38	9,50	60	73	7,00	25	108	9,50	60
4	8,50	45	39	7,00	25	74	8,50	45	109	4,20	20
5	9,50	55	40	11,00	65	75	9,50	61	110	9,50	60
6	8,50	50	41	9,50	55	76	7,00	30	111	11,00	65
7	7,00	30	42	11,00	65	77	8,50	50	112	8,50	50
8	9,50	60	43	9,50	62	78	8,50	45	113	8,50	55
9	9,50	56	44	9,50	56	79	9,50	60	114	8,50	50
10	8,50	45	45	7,00	35	80	11,00	65	115	7,00	25
11	13,80	65	46	8,50	40	81	8,50	40	116	9,50	62
12	8,50	45	47	9,50	60	82	13,80	65	117	9,50	56
13	7,00	30	48	7,00	40	83	7,00	35	118	9,50	60
14	9,50	65	49	8,50	55	84	8,50	50	119	8,50	40
15	9,50	57	50	8,50	50	85	8,50	45	120	11,00	65
16	9,50	60	51	7,00	35	86	8,50	55	121	8,50	55
17	8,50	30	52	8,50	45	87	8,50	45	122	7,00	30
18	11,00	65	53	7,00	35	88	9,50	56	123	8,50	45
19	7,00	35	54	9,50	55	89	11,00	65	124	8,50	50
20	8,50	56	55	8,50	50	90	7,00	36	125	8,50	50
21	8,50	50	56	7,00	30	91	8,50	55	126	11,00	65
22	7,00	35	57	9,50	61	92	9,50	66	127	9,50	55
23	9,50	65	58	8,50	45	93	8,50	50	128	7,00	35
24	9,50	60	59	11,00	66	94	11,00	65	129	8,50	50
25	8,40	45	60	9,50	65	95	9,50	55	130	7,00	35
26	8,50	50	61	11,00	65	96	9,50	60	131	9,50	60
27	11,00	65	62	8,50	55	97	11,00	65	132	11,00	65
28	9,50	65	63	8,50	45	98	8,50	50	133	8,50	55
29	11,00	65	64	8,50	50	99	11,00	65	134	8,50	50
30	8,50	46	65	9,50	55	100	8,50	50	135	9,50	60
31	11,00	65	66	11,00	65	101	9,50	60	136	8,50	57
32	11,00	65	67	8,50	50	102	7,00	37	137	8,50	55
33	9,50	60	68	9,50	65	103	9,50	55	138	7,00	25
34	9,50	65	69	7,00	30	104	11,00	67	139	11,00	68
35	7,00	35	70	8,50	45	105	8,50	51	140	8,50	50

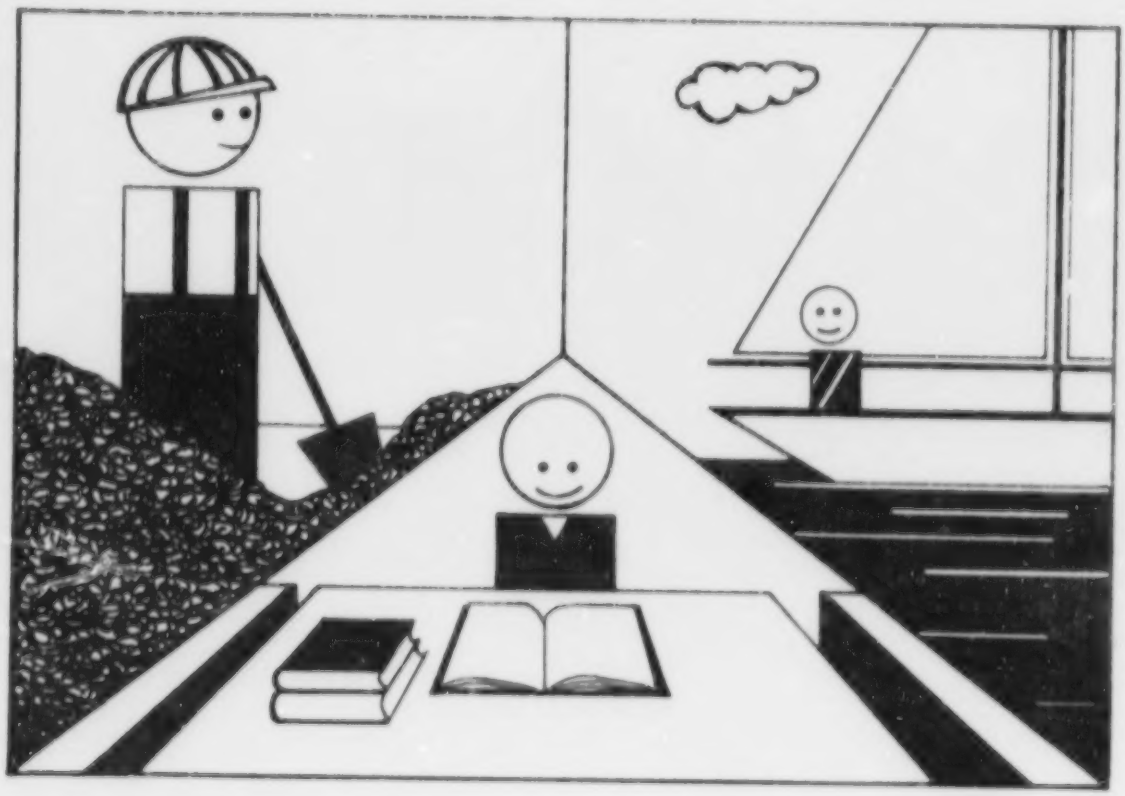
Tableau 3 (suite) Le salaire horaire et l'âge des employés(es) des Mines Leclerc inc.

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
141	9,50	55	169	8,50	50	197	8,50	46	225	8,50	45
142	8,50	40	170	11,00	66	198	7,00	35	226	9,50	55
143	11,00	65	171	8,50	50	199	9,50	55	227	9,50	60
144	9,50	60	172	9,50	55	200	9,50	56	228	8,50	45
145	9,50	60	173	9,50	60	201	9,50	61	229	8,50	50
146	13,80	65	174	8,50	40	202	9,50	60	230	9,50	60
147	8,50	45	175	4,20	20	203	9,50	60	231	8,50	50
148	8,40	40	176	9,50	56	204	11,00	65	232	4,20	20
149	9,50	55	177	7,00	30	205	7,00	25	233	9,50	55
150	9,50	61	178	9,50	60	206	8,50	50	234	11,00	65
151	7,00	30	179	8,50	55	207	9,50	65	235	8,50	50
152	8,50	40	180	8,50	45	208	9,50	60	236	9,50	55
153	7,00	25	181	9,50	55	209	8,50	50	237	11,00	65
154	8,50	45	182	8,50	50	210	7,00	35	238	9,50	60
155	11,00	66	183	11,00	65	211	9,50	65	239	11,00	65
156	7,00	30	184	8,50	40	212	8,50	40	240	9,50	56
157	8,50	40	185	13,80	65	213	9,50	55	241	9,50	65
158	11,00	67	186	9,50	55	214	9,50	60	242	7,00	35
159	9,50	60	187	9,50	60	215	8,50	45	243	8,50	45
160	8,50	40	188	11,00	65	216	7,00	40	244	8,50	51
161	8,50	45	189	8,50	40	217	9,50	55	245	9,50	60
162	8,50	41	190	4,20	21	218	9,50	60	246	7,00	30
163	9,50	56	191	8,50	40	219	8,50	50	247	8,50	40
164	9,50	60	192	11,00	66	220	8,50	40	248	9,50	60
165	9,50	60	193	9,50	55	221	8,50	45	249	9,50	60
166	4,20	20	194	9,50	60	222	13,80	65	250	8,50	45
167	9,50	60	195	9,50	60	223	9,50	60			
168	11,00	65	196	11,00	65	224	8,50	50			

Tableau 4 Le salaire horaire et l'âge des employés(es) de La Boulangerie Béland

No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge	No	Taux horaire	Âge
1	11,00	30	31	7,80	45	61	11,00	30	91	9,90	35	121	9,90	51
2	7,80	50	32	12,10	60	62	7,80	50	92	8,90	52	122	9,90	40
3	12,10	35	33	11,00	27	63	12,10	45	93	12,10	45	123	12,10	51
4	12,10	65	34	7,80	55	64	8,90	50	94	9,90	40	124	11,00	40
5	8,90	25	35	12,10	60	65	11,00	35	95	12,10	40	125	9,90	45
6	11,00	30	36	8,90	44	66	7,80	55	96	7,80	56	126	8,90	30
7	8,90	45	37	8,90	35	67	12,10	35	97	12,10	25	127	12,10	40
8	12,10	51	38	12,10	60	68	8,90	65	98	8,90	60	128	11,00	50
9	7,80	35	39	12,10	40	69	12,10	30	99	12,10	40	129	12,10	45
10	11,00	55	40	9,90	35	70	8,90	56	100	11,00	45	130	9,90	30
11	12,10	20	41	11,00	50	71	8,90	30	101	7,80	25	131	9,90	30
12	11,00	30	42	8,90	41	72	11,00	45	102	11,00	62	132	7,80	41
13	7,80	41	43	11,00	25	73	7,80	57	103	12,10	47	133	9,90	55
14	8,90	52	44	7,80	40	74	11,00	35	104	11,00	30	134	11,00	40
15	9,90	35	45	12,10	65	75	9,90	50	105	8,90	45	135	12,10	20
16	9,90	35	46	11,00	35	76	12,10	25	106	12,10	57	136	8,90	45
17	11,00	40	47	8,90	45	77	7,80	55	107	8,90	35	137	9,90	35
18	8,90	46	48	12,10	25	78	12,10	40	108	12,10	45	138	12,10	45
19	12,10	20	49	11,00	47	79	8,90	55	109	11,00	55	139	12,10	50
20	12,10	31	50	7,80	40	80	11,00	20	110	7,80	35	140	7,80	26
21	9,90	42	51	7,80	35	81	9,90	50	111	12,10	55	141	9,90	65
22	7,80	55	52	9,90	50	82	7,80	40	112	11,00	25	142	11,00	35
23	8,90	35	53	9,90	40	83	12,10	50	113	11,00	66	143	9,90	46
24	11,00	45	54	12,10	46	84	8,90	31	114	8,90	30	144	8,90	50
25	7,80	21	55	9,90	30	85	11,00	53	115	9,90	50	145	9,90	30
26	12,10	30	56	8,90	35	86	9,90	40	116	12,10	45	146	12,10	45
27	11,00	40	57	12,10	45	87	9,90	60	117	12,10	40	147	12,10	25
28	9,90	45	58	12,10	36	88	9,90	32	118	9,90	60	148	9,90	50
29	9,90	35	59	11,00	61	89	11,00	51	119	11,00	40	149	9,90	41
30	11,00	56	60	11,00	46	90	8,90	60	120	11,00	60	150	11,00	45

Les activités étudiantes du groupe au cours des dernières vacances estivales



APPLICATION II

NOM(S): _____

_____

_____

OBJECTIF GÉNÉRAL

Acquérir les concepts statistiques et développer les habiletés nécessaires à l'analyse des données.

MATÉRIEL

Un questionnaire

CONTEXTE DE L'APPLICATION

Nous voulons étudier les activités que les étudiants(es) du groupe ont pratiquées au cours des dernières vacances estivales. Cette étude fait appel à la statistique descriptive. Nous proposons également une méthodologie pour faire cette étude.

CHEMINEMENT

Présentation du problème

Nous voulons de l'information sur les activités étudiantes des dernières vacances estivales de notre groupe. Des questions nous viennent à l'esprit:

Quel genre d'activités ont exercé les étudiants(es) du groupe durant l'été?

Quel temps a été consacré aux différentes activités?

Est-ce que le milieu des étudiants(es) influence le genre d'activités qu'ils(elles) ont exercé durant l'été?

Quels(les) sont ceux(celles) qui ont travaillé et quel genre de travail ont-ils(elles) occupé?

Quels(les) sont ceux(celles) qui ont étudié et quel genre d'études ont-ils(elles) suivies?

Quels(les) sont ceux(celles) qui ont voyagé et quel genre de voyages ont-ils(elles) faits?

Quels(les) sont ceux(celles) qui ont exercé d'autres activités et quelles sont ces autres activités?

Quels sont les revenus des étudiants(es) du groupe durant l'été?

Quelles ont été les dépenses encourues pour ces différentes activités?

Puisque le nombre d'étudiants(es) n'est pas trop élevé, nous les contacterons tous.

Méthodologie

Afin de solutionner le problème posé, il faut se donner une marche à suivre. Dans le but de recueillir l'information, nous utiliserons un questionnaire portant sur les différentes activités étudiantes. Nous vous présenterons un exemple de questionnaire.

La première partie du questionnaire a pour but d'avoir des renseignements sociologiques sur les étudiants(es) tandis que les autres parties vont nous permettre de connaître leurs activités estivales.

Généralement, on doit restreindre le nombre de questions à cause de la difficulté à recueillir l'information et surtout à l'analyser Le questionnaire doit être conçu en fonction du dépouillement et de l'analyse qui en seront faits. Ainsi, les questions fermées sont plus faciles à dépouiller que les questions ouvertes.

Cueillette des données

Lorsque le questionnaire est bâti, il faut aller chercher l'information. Pour cela, chaque étudiant(e) du groupe remplit le questionnaire.

Remplir le questionnaire, si ce n'est pas déjà fait.

Par la suite, chaque étudiant(e) aura un exemplaire de l'ensemble des réponses obtenues auprès du groupe.

Tableau d'observations et codification

Maintenant que chaque étudiant(e) possède l'ensemble de l'information obtenue des questionnaires, il(elle) peut commencer à se pencher sur les questions posées au début et essayer d'y répondre.

Pour répondre à ces questions, il faut dépouiller l'ensemble des questionnaires. Pour éviter de revenir continuellement aux questionnaires, on te conseille de faire un tableau d'observations, qui recueillera toutes les données utiles à chaque élève, comme tu en as fait un à la première expérience du laboratoire expérimental.

Ce tableau est assez facile à construire, mais avant de le compléter, il te faudra établir une certaine règle de procédure, qu'on appelle codification, pour te simplifier l'entrée des données. Par exemple, pour la caractéristique sexe, au lieu d'écrire les catégories: masculin et féminin, tu peux facilement convenir d'écrire M, F; au lieu d'écrire les catégories: avion, train, bateau, automobile, autobus, "pouce", motocyclette, bicyclette pour la caractéristique moyen de transport, tu peux décider d'employer les lettres dans l'ordre: A, B, C, D, E, E, G, H, ou les chiffres 1 à 8.

Donc, avant de te pencher sur l'analyse, tu dois bâtir un tableau d'observations en prenant la peine de codifier d'abord les catégories.

Codifier les réponses du questionnaire.

CODIFICATION

QUESTION	CARACTÉRISTIQUE	CATÉGORIE	CODIFICATION
1.1	Sexe	Masculin	M
		Féminin	F
1.2	Niveau	Collège I	1
		Collège II	2
		Collège III	3
1.3	Secteur	Général	
		Professionnel	
1.4	Orientation	3 chiffres	
1.5	Âge		
1.6	Date de naissance		
1.7	Résidence	Urbain	
		Rural	
1.8	Profession du père		
1.9	Profession de la mère		

QUESTION	CARACTÉRISTIQUE	CATÉGORIE	CODIFICATION
2.1	Travail	Oui Non	
2.2	Découverte du travail	Bureau de placement Emploi de l'année précédente Journaux Famille Amis(es) Autres	
2.3	Type de travail	Manuel Intellectuel	
2.4	Connexité aux études	Pas du tout Un peu Beaucoup Complètement	
2.5	Milieu de travail	Secteur public Secteur privé Famille Chez des amis(es) Autres	
2.6	Distance	Moins de 3 km 3 à 10 km 10 à 50 km Plus de 50 km	
2.7	Nombre de semaines		
2.8	Nombre de jours/ semaine		
2.9	Nombre d'heures/jour		

QUESTION	CARACTÉRISTIQUE	CATÉGORIE	CODIFICATION
2.10	Salaire	Oui Non	
2.11	Montant du gain		
3.1	Études	Oui Non	
3.2	Institution	- Collège de Rimouski - Collège de Rivière-du-Loup - - -	
3.3	Pourquoi les études	- Finir mon D.E.C. - M'avancer - - -	
3.4	Nombre de semaines		
3.5	Nombre de jours/semaine		
3.6	Nombre d'heures/jour		
3.7	Moment de l'étude	Jour Soir	
3.8	Matière étudiée		
3.9	Coût total $		
4.1	Voyages	Oui Non	

QUESTION	CARACTÉRISTIQUE	CATÉGORIE	CODIFICATION
4.2	Lieu des voyages	Québec Autres provinces Etats-Unis Ailleurs	
4.3	Distance		
4.4	Moyen de transport	Avion Train Bateau Automobile Autobus " Pouce " Motocyclette Bicyclette Autres	
4.5	Jours de voyage		
4.6	Coût des voyages		
5.1	Autres activités	- Travail individuel - Sport - Travail de maison - Vagabondage - Camping - Sorties - Musique - - - - -	

QUESTION	CARACTÉRISTIQUE	CATÉGORIE	CODIFICATION
5.2	Nombre d'heures		
5.3	Dépenses		
6.1	Satisfaction de tes vacances	Pas du tout Très peu Assez Beaucoup	

Remarques sur la codification

Certaines questions ouvertes donnent possiblement plusieurs réponses, on te conseille d'essayer de regrouper ces diverses réponses.

Pour les questions qui ne demandent pas de codification, mais une réponse chiffrée, on te demande de faire attention à la dimension maximale de ces nombres, soit en construisant le tableau d'observations ou en le complétant s'il est déjà construit.

Remplir le tableau d'observations.

Analyse statistique des données

A l'étape précédente, les informations au sujet des activités estivales des étudiants(es) ont été regroupées dans un tableau d'observations; il s'agit maintenant d'analyser ces informations et d'en extraire l'essentiel.

Il paraît logique de se donner d'abord une idée générale des activités exercées par les étudiants(es) avant d'aborder d'une façon détaillée chaque activité.

Ainsi, dans un premier temps, nous étudierons le genre, le temps consacré et les dépenses encourues par les étudiants(es) pour les activités. Dans un deuxième temps, nous regarderons les activités: le travail, les voyages, les études, en cherchant leur genre et quelles personnes les ont exercées. Il en est de même des autres activités mentionnées dans l'enquête. Pour l'activité travail, on devra cependant ajouter la dimension salaire.

Techniquement, l'analyse de l'enquête peut se faire question par question. D'ailleurs, au laboratoire expérimental, nous nous sommes familiarisés à des outils servant à l'analyse d'une caractéristique: tableau d'effectifs, représentation graphique, moyenne arithmétique, mode, médiane, étendue et écart moyen. Ces outils permettent de découvrir et de présenter visuellement la structure d'une caractéristique et de la résumer en cherchant sa tendance générale et sa dispersion.

Une telle façon de procéder nous éloigne un peu de nos préoccupations, de la raison d'être de notre enquête, des thèmes

A. 2. 13 et A 2. 14 manquents

Analyse statistique des données

A l'étape précédente, les informations au sujet des activités estivales des étudiants(es) ont été regroupées dans un tableau d'observations; il s'agit maintenant d'analyser ces informations et d'en extraire l'essentiel.

Il paraît logique de se donner d'abord une idée générale des activités exercées par les étudiants(es) avant d'aborder d'une façon détaillée chaque activité.

Ainsi, dans un premier temps, nous étudierons le genre, le temps consacré et les dépenses encourues par les étudiants(es) pour les activités. Dans un deuxième temps, nous regarderons les activités: le travail, les voyages, les études, en cherchant leur genre et quelles personnes les ont exercées. Il en est de même des autres activités mentionnées dans l'enquête. Pour l'activité travail, on devra cependant ajouter la dimension salaire.

Techniquement, l'analyse de l'enquête peut se faire question par question. D'ailleurs, au laboratoire expérimental, nous nous sommes familiarisés à des outils servant à l'analyse d'une caractéristique: tableau d'effectifs, représentation graphique, moyenne arithmétique, mode, médiane, étendue et écart moyen. Ces outils permettent de découvrir et de présenter visuellement la structure d'une caractéristique et de la résumer en cherchant sa tendance générale et sa dispersion.

Une telle façon de procéder nous éloigne un peu de nos préoccupations, de la raison d'être de notre enquête, des thèmes

que nous désirons étudier. Il est préférable de disséquer l'enquête en travaillant par thème et protocole. En fait, on cerne un sujet et on essaie d'y répondre en choisissant les questions pertinentes.

Les principaux thèmes de l'enquête et leurs protocoles:

1- Les différentes activités
   a) Genre d'activités
   b) Temps consacré aux activités
   c) Dépenses pour les différentes activités

2- Le travail
   a) Qui a travaillé?
   b) Genre de travail
   c) Salaire

3- L'étude
   a) Qui a étudié?
   b) Genre d'étude

4- Les voyages
   a) Qui a voyagé?
   b) Genre de voyages

5- Autres activités
   a) Qui a exercé ces activités?
   b) Genre d'activités.

Commençons l'analyse du thème, les différentes activités, par le protocole, genre d'activités.

Quelles sont les questions de l'enquête qui se rapportent à ce protocole?

A première vue, les questions 2.1, 3.1, 4.1 et 5.1. Ces questions nous permettent de dégager le pourcentage d'étudiants(es) qui ont exercé chacune de ces activités.

Compléter le tableau.

Tableau 1 Le pourcentage d'étudiants(es) qui ont exercé les activités: travail, études, voyages et autres

ACTIVITÉ	POURCENTAGE (%)
Travail	----------
Études	----------
Voyages	----------
----------	----------
----------	----------
----------	----------

On peut visualiser le pourcentage d'étudiants(es) qui ont exercé les différentes activités en faisant une représentation graphique.

Compléter cette représentation graphique.

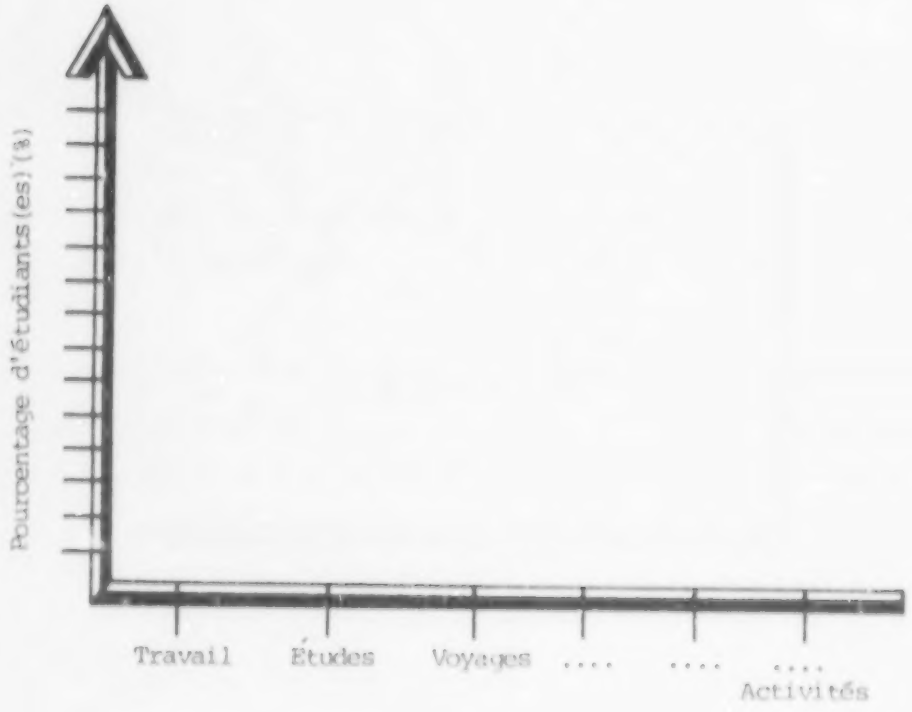


Figure 1 Le pourcentage d'étudiants(es) qui ont exercé les activités: travail, études, voyages et .....

Pour toutes difficultés relatives aux représentations graphiques, voir le manuel d'instructions à ce sujet.

En complétant le tableau 1, on a constaté que des étudiants(es) ont exercé plus d'une activité. Quelques étudiants(es) ont travaillé et voyagé, d'autres ont étudié et voyagé, etc...

On peut représenter facilement les idées à l'aide d'un diagramme de Venn.

Complétion du diagramme

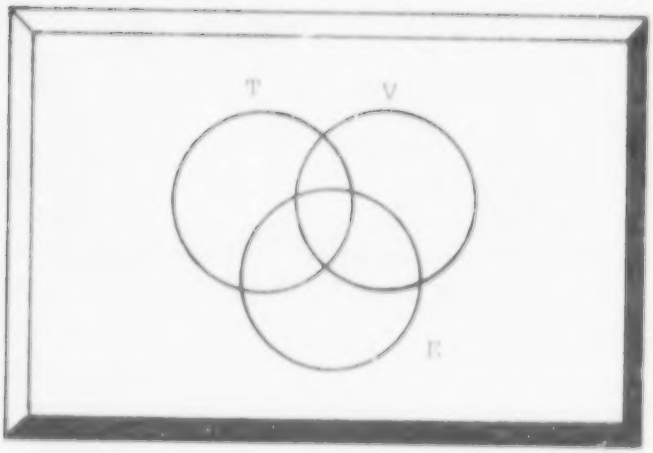


Figure 2 Le pourcentage d'étudiants(es) qui ont exercé une, deux ou trois activités (travail (T), voyages (V), études (E))

Trouver un autre moyen de représenter ce phénomène.

Interpréter les résultats obtenus jusqu'à maintenant (Tableau 1, figure 1, figure 2...).

Pour compléter l'étude du thème "les différentes activités", il nous reste à étudier deux protocoles:

1) temps consacré aux activités,
2) dépenses pour les différentes activités.

Ces protocoles seront étudiés par une équipe d'étudiants(es) qui devront présenter les résultats de leur analyse aux autres étudiants(es) d'une façon verbale et écrite. Il en est de même des autres protocoles sauf celui du salaire qui devra être fait par chaque étudiant(e).

La classe est divisée en cinq équipes et les protocoles sont attribués d'une façon aléatoire à chaque équipe.

Équipe	Thème	Protocole
1	Les différentes activités	- Temps consacré aux activités - Dépenses pour les différentes activités
2	Le travail	- Qui a travaillé? - Genre de travail
3	L'étude	- Qui a étudié? - Genre d'étude
4	Les voyages	- Qui a voyagé? - Genre de voyage
5	Les autres activités	- Qui a exercé ces activités? - Genre d'activités

Le salaire des étudiants(es)

Plusieurs questions de l'enquête se rapportent au salaire des étudiants(es), particulièrement les questions 2.1, 2.7, 2.8, 2.9, 2.10, 2.11.

Que peut-on en sortir de significatif?

Il serait intéressant de connaître le revenu gagné par les étudiants(es). Le montant moyen du revenu et sa dispersion. Existe-t-il une relation entre l'âge des étudiants(es) et leur revenu, et si oui, quelle est cette relation?

De plus, il serait intéressant de voir comment le revenu des étudiants(es) se répartit selon le sexe, l'âge, le niveau scolaire, le secteur, la profession des parents.

Également, il faudrait essayer de connaître le salaire horaire des étudiants(es) pour vérifier s'il est plus élevé ou plus faible que le salaire minimum.

Pour répondre à ces différentes questions, on doit se servir des outils de notre coffre.

Le revenu gagné par les étudiants(es)

Faire un tableau de fréquences du revenu gagné par les étudiants(es).

Tableau **Le revenu gagné par les étudiants(es) pendant leurs dernières vacances**

Montant ($)	Effectif	Fréquence

Interpréter ces résultats.

Évaluer la tendance centrale et la dispersion du revenu des étudiant(e)s.

Tendance centrale

1) Moyenne arithmétique _____

2) Mode ..... _____

3) Médiane ..... _____

Dispersion

1) Étendue ..... _____

2) Écart moyen ..... _____

3. Variance ..... _____

Interpréter ses résultats.

Il serait intéressant de vérifier s'il existe une relation entre l'âge des étudiants(es) et leur revenu.

Remplir le tableau suivant.

Tableau__ L'âge des étudiants(es) et leur revenu

Numéro de questionnaire	Âge	Revenu

Pour vérifier cette relation, il existe différents moyens.

1) Moyen visuel

Faire une représentation graphique entre l'âge et le revenu de l'étudiant(e).

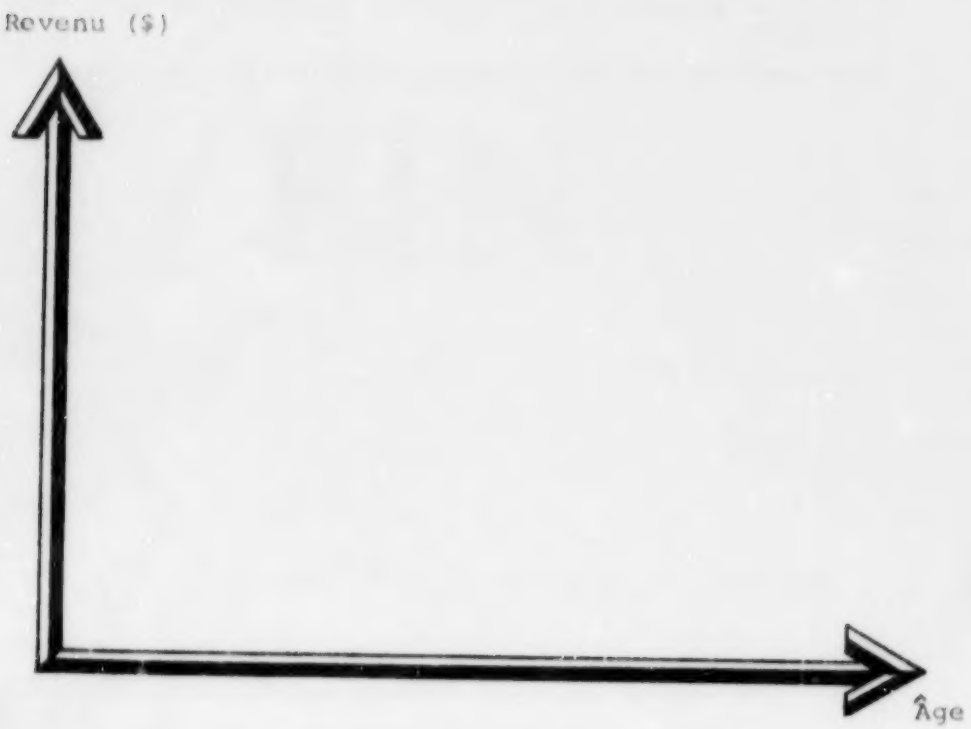


Figure Le revenu des étudiants(es) en fonction de leur âge

2) Moyen analytique

Calculer le coefficient de corrélation linéaire.

r = _____

Voir le manuel d'instructions au sujet de cet instrument.

Exprimer la relation si elle existe.

Revenu des étudiants(es) = ? âge des

des étudiants(es) + ?

Voir le manuel d'instructions.

Interpréter ces résultats.

Étude du revenu étudiant selon le sexe

Pour analyser le revenu selon le sexe, on peut faire un tableau de fréquences, une représentation graphique et calculer les mesures de tendance centrale et de dispersion pour les étudiantes et faire de même pour les étudiants.

Faire un tableau de fréquences du revenu gagné par les étudiantes.

Tableau... Le revenu gagné par les étudiantes pendant les dernières vacances estivales

Revenu ($)	Effectif	Fréquence

Faire une représentation graphique du revenu gagné par les étudiantes.

Figure Le revenu gagné par les étudiantes pendant les dernières vacances estivales

Évaluer la tendance centrale et la dispersion du revenu gagné par les étudiantes.

Médiane ..... _____

Moyenne arithmétique _____

Mode ..... _____

Étendue ..... _____

Écart-type ..... _____

On fait de même pour les étudiants de sexe masculin.

Faire un tableau de fréquences du revenu gagné par les étudiants.

Tableau ... Le revenu gagné par les étudiants pendant les dernières vacances estivales

Revenu ($)	Effectif	Fréquence

Faire une représentation graphique du revenu gagné par les étudiants.

Figure Le revenu gagné par les étudiants pendant les dernières vacances estivales

Évaluer la tendance centrale et la dispersion du revenu gagné par les étudiants.

Moyenne arithmétique _____

Mode ..... _____

Médiane ..... _____

Étendue..... _____

Écart-type..... _____

Pour comparer le revenu des étudiants à celui des étudiantes, on a intérêt à présenter l'information dans un seul tableau ou une seule représentation graphique.

Faire un tableau pour représenter le revenu gagné par les étudiants et les étudiantes.

Tableau... Le revenu gagné par les étudiants et les étudiantes pendant les dernières vacances estivales

Revenu ($)	Fréquence	
	Masculin	Féminin
Total	100%	100%

Faire une représentation graphique du revenu gagné par les étudiants et les étudiantes.

Compléter le tableau suivant

Tableau ... La tendance centrale et la dispersion du revenu des étudiants(es) selon le sexe

Sexe	Tendance centrale			Dispersion	
	Moyenne arithmétique	Mode	Médiane	Étendue	Écart-type
Féminin	_____	_____	_____	_____	_____
Masculin	_____	_____	_____	_____	_____

Interpréter ces résultats.

Étude du revenu étudiant selon le niveau scolaire

On pourrait comme l'étude du revenu étudiant selon le sexe, faire un tableau de fréquences, une représentation graphique pour chaque catégorie du niveau scolaire. Toutefois, comme on l'a vu, il n'y a pas d'intérêt pratique à travailler ainsi. Il est préférable lorsque c'est possible, de regrouper les informations dans un seul tableau ou dans une seule représentation graphique. Les comparaisons sont alors plus faciles à faire.

Faire un tableau de fréquence du revenu gagné par les étudiants(es) de collégial I, collégial II et collégial III.

Tableau... Le revenu gagné par les étudiants(es) selon le niveau scolaire

Revenu ($)	Niveau scolaire		
	collégial I	collégial II	collégial III

Faire une représentation graphique du revenu gagné par les étudiants(es) selon le niveau scolaire.

Figure Le revenu gagné par les étudiants(es) selon le niveau scolaire

Compléter le tableau suivant.

Tableau... La tendance et la dispersion du revenu des étudiants(es) selon le niveau scolaire

	Tendance centrale			Dispersion	
Niveau scolaire	Moyenne arithmétique	Mode	Médiane	Étendue	Écart-type
Collégial I	_____	_____	_____	_____	_____
Collégial II	_____	_____	_____	_____	_____
Collégial III	_____	_____	_____	_____	_____

Interpréter ces résultats.

Étude du revenu étudiant selon le secteur

Faire un tableau de fréquences du revenu gagné par les étudiants(es) du secteur professionnel et du secteur général.

Tableau... Le revenu gagné par les étudiants(es) selon le secteur

Revenu ($)	Secteur	
	Professionnel	Général

Faire une représentation graphique du revenu des étudiants(es) selon le secteur

Figure .... Le revenu gagné par les étudiants(es) selon le secteur

Compléter le tableau suivant.

Tableau....La tendance centrale et la dispersion du revenu des étudiants(es) selon le secteur

	Tendance centrale			Dispersion	
Secteur	Moyenne arithmétique	Mode	Médiane	Étendue	Écart-type
Professionnel	_____	_____	_____	_____	_____
Général	_____	_____	_____	_____	_____

Interpréter ces résultats.

Étude du revenu étudiant selon l'âge

Faire un tableau de fréquences : revenu gagné par les étudiants(es) selon leur âge.

Tableau... Le revenu gagné par les étudiants(es) selon l'âge

	Âge					
Revenu ($)						
	.....	.....	.....	.....	.....	.....

Faire une représentation graphique du revenu gagné par les étudiants(es) selon l'âge.

Compléter le tableau suivant.

Tableau... La tendance centrale et la dispersion du revenu des étudiants(es) selon l'âge

	Tendance centrale			Dispersion	
Âge	Moyenne arithmétique	Mode	Médiane	Étendue	Écart-type

Interpréter ces résultats.

Étude du revenu étudiant selon la profession du père et de la mère

Faire un tableau de fréquences ou une représentation graphique du revenu gagné par les étudiant(e)s selon la profession du père.

Interpréter ces résultats.

Faire un tableau de fréquences ou une représentation graphique du revenu gagné par les étudiants(es) selon la profession de la mère.

Interpréter ces résultats.

Le salaire horaire des étudiants(es)

Même si nous avons demandé le salaire gagné en tout, il est possible de connaître le salaire horaire de chaque étudiant(e). En effet, dans notre enquête, nous avons demandé le nombre de semaines travaillées, le nombre de jours par semaine et le nombre d'heures par jour.

Calculer le salaire hebdomadaire de chaque étudiant(e) et remplir le tableau.

Tableau... Le salaire horaire de chaque étudiant(e) selon le sexe

Numéro du questionnaire	Salaire horaire ($)	Sexe

Il serait intéressant de comparer la tendance centrale et la dispersion du salaire horaire des étudiants à celles des étudiantes.

Remplir le tableau.

Tableau... La tendance centrale et la dispersion du salaire horaire des étudiants(es) selon le sexe

	Étudiant(e)		
	Masculin	Féminin	Ensemble
Moyenne arithmétique			
Écart-type			

Interpréter ces résultats.

Nous avons présenté surtout l'aspect technique de l'analyse statistique. Cependant, dans un rapport, il est essentiel d'ajouter de la chair à ce squelette, c'est-à-dire de soulever, de préciser les figures et les tableaux introduits dans la section.

Interprétation

Cette partie de la méthodologie consiste essentiellement à revenir sur les résultats obtenus à l'analyse statistique, à les discuter, à faire ressortir les éléments principaux, les faits intéressants, à les expliquer et interpréter dans le contexte de l'étude.

Interpréter les résultats des protocoles, genre d'activités et salaire des étudiants(es).

Conclusion

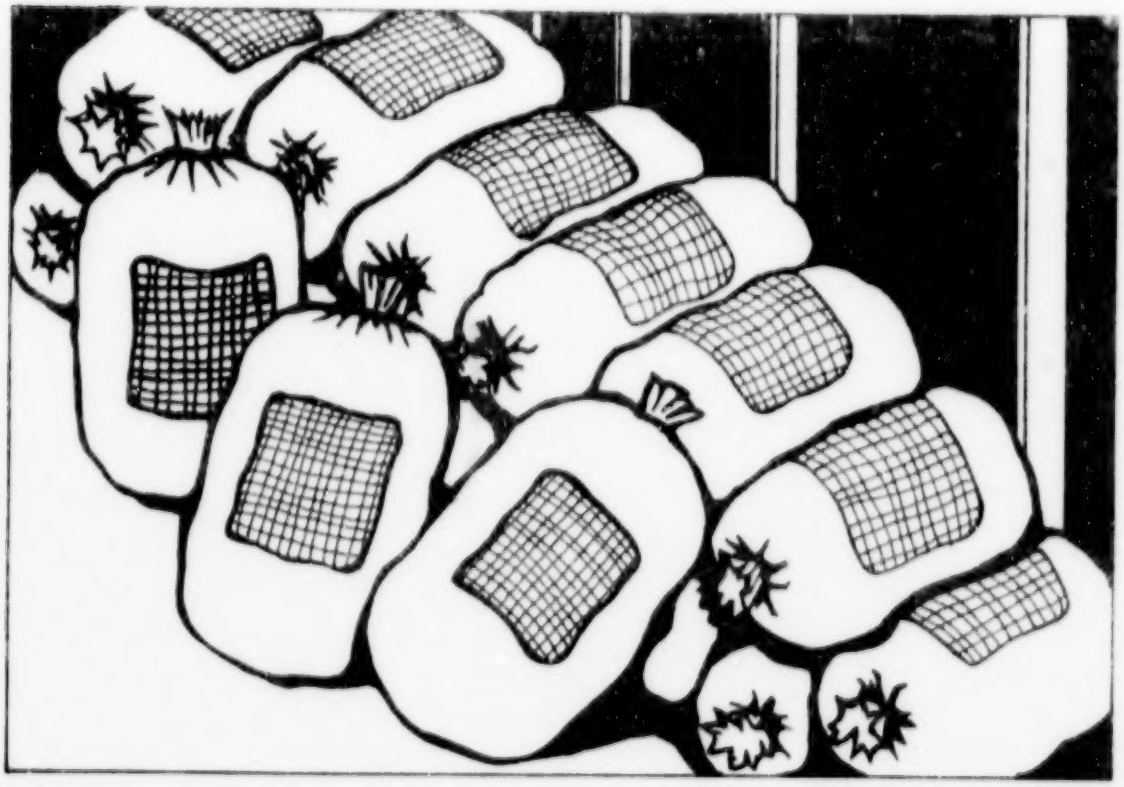
Dans la conclusion, on doit ordinairement rappeler le sujet étudié, la méthode employée et les principaux résultats obtenus.

Rédiger la conclusion de cette enquête en ne considérant que les deux protocoles, genre d'activités et valeur des étudiants(es).

En respectant la méthode proposée, rédiger un rapport final sur les protocoles, genre d'activités et salaires des étudiant(e)s.

2. Théorie d'échantillonnage

Étude des plaintes formulées par les clients de la coopérative d'alimentation de Rimouski sur la qualité des pommes de terre



APPLICATION III

Nom(s) _____

_____

OBJECTIF GÉNÉRAL

Acquérir les concepts statistiques et développer les habiletés nécessaires à l'analyse des données.

MATÉRIEL

Aucun.

CONTEXTE DE L'APPLICATION

Nous voulons étudier les plaintes formulées par les clients de la Coopérative d'alimentation de Rimouski, sur la qualité des pommes de terre. Cette étude fait appel à la théorie d'échantillonnage.

Depuis quelques semaines, les clients de la coopérative d'alimentation de Rimouski se plaignent auprès du gérant de la mauvaise qualité des pommes de terre.

Sensible aux plaintes des consommateurs, le gérant décide de vérifier la qualité de la dernière livraison et fait ouvrir cinq (5) sacs de pommes de terre. Deux sacs contiennent de nombreuses pommes de terre pourries et curieusement, tous les sacs pèsent moins de 4540 g, poids indiqué sur le sac (ou 4,54 kg).

Q 1. Suite à ces constatations, le gérant devrait-il refuser les 4 000 sacs reçus?

Expliquer.

Le contrat de vente avec le producteur mentionne entre autres que le poids d'un sac doit être autour de 4540 g et ne jamais être inférieur à 4400 g et que le pourcentage de sacs pourris ne peut dépasser 10% du nombre total.

Q 2. *Crois-tu que la méthode employée par le gérant permet d'évaluer la qualité des 4 000 sacs de pommes de terre? Le nombre de sacs vérifiés est-il suffisant selon toi?*

Commenter et proposer une méthode.

La statistique offre quelques méthodes pour solutionner ce problème.

(Voir le manuel d'instruc[illegible]

La méthode la plus simple consiste à prendre au hasard une quantité suffisante de sacs parmi les 4000 sacs.

Le gérant ayant été sensibilisé aux méthodes statistiques décide, suite à ses premières observations, de les employer pour vérifier si les deux principales clauses du contrat sont respectées. Le poids moyen des sacs de pommes de terre est-il vraiment inférieur au poids indiqué et le pourcentage de sacs pourris dépasse-t-il réellement les 10%?

Construire un échantillon de 10 sacs pris au hasard pour faire les tests.

1. Base de sondage

2. *Lire la table de nombres au hasard.*
(1ère page ...)

3. *L'échantillon*

4. *Test du poids des sacs*

5. *Test du pourcentage de sacs pourris*

Q 8. *À partir des résultats des tests, le grossiste a-t-il raison de douter du respect du contrat?*

Expliquer.

Cependant, avant d'aller rencontrer le producteur pour discuter des clauses non respectées du contrat, le gérant veut avoir une bonne idée du poids moyen réel des sacs et du pourcentage de sacs pourris. L'erreur d'estimation ne devrait pas, selon lui, dépasser 10 grammes pour le poids moyen et 7% pour le pourcentage de sacs pourris.

Déterminer la quantité de sacs à choisir au hasard pour respecter l'erreur d'estimation, jugée utile par le gérant, sur le poids moyen des sacs et le pourcentage de sacs pourris, à un niveau de confiance de l'ordre de 95%.

1. *Le poids moyen des sacs*

   Erreur acceptable : $\Delta p_1 = \pi_{.}\ \pi_{.}$

   $n = ?$

2. *Le pourcentage de sacs pourris*

   Erreur acceptable : $\pi_{.}\ \pi_{.} = \pi_{.}\ \pi_{.}$

   $n = ?$

Construire l'échantillon.

(Les sacs sont disposés dans l'entrepôt en douze rangées. Leurs poids sont donnés aux pages suivantes et certains sont soulignés à cause de la pourriture.)

1. *Base de sondage*

2. *Lire la table de nombres au hasard.*
(1ère colonne ici)

3. L'échantillon

4. L'estimé du poids moyen des eaux

5. L'estimé du pourcentage de eaux pourris

No	1-	rangée	(26x13)		Poids	(g)		L'entrepôt					
1	4470	4470	4505	4505	4505	4540	4540	4505	4470	4505	4470	4400	4435
14	4505	4435	4505	4505	4505	4505	4470	4540	4540	4505	4435	4505	4470
27	4505	4505	4505	4505	4505	4505	4470	4575	4435	4470	4575	4505	4540
40	4435	4470	4505	4470	4505	4470	4505	4505	4505	4540	4505	4435	4540
53	4470	4505	4505	4505	4540	4540	4470	4470	4470	4505	4470	4400	4435
66	4505	4540	4470	4470	4505	4505	4610	4435	4470	4540	4505	4505	4540
79	4470	4435	4470	4435	4540	4505	4470	4505	4435	4470	4470	4505	4435
92	4505	4470	4505	4470	4470	4505	4470	4505	4470	4470	4505	4505	4505
105	4505	4505	4400	4470	4505	4505	4470	4575	4470	4470	4435	4540	4470
118	4505	4505	4505	4470	4435	4505	4435	4505	4470	4435	4470	4505	4470
131	4505	4505	4470	4470	4435	4470	4505	4505	4470	4470	4505	4470	4435
144	4470	4435	4610	4435	4470	4505	4505	4470	4505	4540	4505	4505	4400
157	4435	4470	4505	4470	4435	4540	4470	4505	4470	4435	4505	4540	4540
170	4435	4470	4540	4435	4470	4540	4435	4505	4470	4540	4540	4470	4470
183	4435	4540	4470	4575	4470	4505	4470	4470	4505	4575	4470	4505	4470
196	4505	4505	4575	4400	4505	4470	4470	4435	4540	4470	4470	4505	4470
209	4470	4470	4540	4505	4505	4505	4470	4505	4540	4470	4470	4505	4505
222	4470	4470	4470	4540	4505	4505	4470	4435	4470	4505	4505	4505	4470
235	4505	4435	4505	4400	4540	4505	4505	4470	4470	4470	4540	4470	4470
248	4505	4470	4400	4435	4435	4470	4575	4505	4540	4540	4400	4470	4505
261	4470	4470	4470	4470	4505	4435	4435	4435	4470	4540	4435	4505	4505
274	4540	4470	4435	4435	4470	4505	4505	4435	4470	4470	4470	4470	4435
287	4540	4470	4470	4505	4470	4435	4470	4470	4505	4540	4540	4470	4505
300	4505	4470	4470	4540	4435	4505	4470	4505	4470	4505	4505	4505	4435
313	4470	4470	4505	4470	4575	4505	4505	4470	4505	4470	4435	4505	4470
326	4470	4470	4400	4400	4435	4470	4505	4505	4540	4540	4435	4435	4435

2- rangée

No													
339	4505	4540	4470	4505	4470	4505	4505	4470	4470	4505	4470	4435	4470
352	4505	4470	4435	4540	4540	4505	4435	4505	4505	4540	4435	4470	4470
365	4470	4470	4505	4470	4505	4505	4540	4505	4470	4470	4505	4505	4400
378	4470	4505	4505	4470	4480	4435	4505	4470	4470	4505	4470	4470	4470
391	4470	4470	4400	4540	4470	4575	4470	4505	4505	4470	4540	4575	4470
404	4505	4505	4435	4505	4435	4470	4470	4505	4470	4470	4470	4435	4470
417	4470	4435	4435	4470	4470	4470	4505	4400	4505	4470	4470	1505	4505
430	4505	4505	4470	4470	4470	4575	4505	4540	4505	4435	4435	4470	4435
443	4435	4435	4400	4435	4505	4470	4540	4470	4505	4505	4470	4505	4470
456	4505	4435	4470	4435	4505	4505	4470	4540	4470	4470	4435	4470	4540
469	4505	4470	4470	4470	4470	4505	4505	4470	4435	4435	4435	4470	4575
482	4540	4470	4540	4540	4505	4540	4505	4435	4505	4505	4470	4435	4435
495	4505	4505	4470	4540	4470	4505	4435	4505	4365	4470	4470	4470	4540
508	4470	4540	4435	4540	4540	4470	4505	4540	4540	4470	4505	4505	4435
521	4505	4470	4540	4505	4470	4505	4505	4505	4435	4505	4540	4505	4470
534	4470	4505	4470	4470	4470	4505	4435	4470	4575	4575	4540	4505	4470
547	4470	4435	4400	4470	4470	4470	4505	4505	4505	4470	4540	4470	4435
560	4470	4505	4400	4470	4470	4540	4435	4470	4505	4470	4435	4470	4505
573	4505	4505	4470	4540	4470	4470	4470	4470	4470	4470	4540	4470	4575
586	4400	4540	4505	4540	4470	4470	4505	4470	4470	4470	4400	4505	4470
599	4435	4470	4435	4505	4540	4505	4470	4470	4540	4505	4505	4505	4470
612	4540	4435	4540	4505	4505	4505	4400	4540	4505	4575	4470	4505	4435
625	4505	4435	4505	4470	4505	4470	4470	4505	4435	4505	4470	4540	4435
638	4505	4505	4505	4540	4400	4505	4470	4505	4505	4505	4505	4470	4540
651	4505	4575	4505	4575	4470	4540	4470	4470	4505	4470	4505	4540	4505
664	4505	4435	4470	4505	4505	4540	4470	4505	4470	4505	4505	4470	4470

3- rangée

No													
677	4540	4505	4435	4505	4505	4470	4540	4505	4505	4470	4505	4470	4505
690	4435	4470	4470	4435	4470	4505	4435	4505	4505	4470	4470	4505	4470
703	4505	4505	4575	4470	4540	4505	4540	4470	4540	4540	4470	4540	4470
716	4435	4435	4540	4505	4470	4435	4540	4540	4540	4470	4435	4575	4540
729	4470	4470	4540	4505	4505	4505	4470	4505	4435	4505	4470	4505	4400
742	4470	4505	4505	4505	4470	4470	4505	4470	4505	4470	4470	4505	4470
755	4435	4470	4470	4470	4470	4505	4505	4435	4540	4470	4505	4470	4540
768	4540	4470	4435	4505	4505	4505	4470	4505	4540	4470	4505	4505	4505
781	4435	4505	4505	4540	4470	4470	4540	4470	4540	4470	4470	4435	4540
794	4540	4435	4470	4505	4505	4505	4540	4540	4470	4540	4505	4435	4470
807	4505	4505	4470	4505	4505	4470	4505	4470	4505	4575	4540	4470	4435
820	4505	4505	4470	4505	4435	4505	4505	4505	4540	4470	4470	4505	4505
833	4505	4575	4505	4435	4435	4470	4540	4470	4470	4470	4505	4610	4470
846	4470	4540	4505	4400	4505	4505	4470	4400	4540	4505	4505	4540	4435
859	4505	4505	4505	4470	4540	4505	4505	4505	4470	4470	4365	4505	4505
872	4505	4435	4470	4470	4505	4470	4435	4540	4470	4505	4505	4470	4470
885	4470	4435	4435	4505	4470	4470	4505	4505	4470	4435	4470	4470	4540
898	4505	4540	4470	4505	4505	4505	4505	4470	4505	4470	4470	4505	4470
911	4435	4435	4470	4470	4470	4470	4470	4505	4505	4470	4470	4540	4505
924	4470	4505	4470	4505	4505	4540	4505	4505	4505	4505	4055	4505	4505
937	4505	4540	4575	4505	4505	4470	4470	4505	4435	4470	4505	4470	4505
950	4505	4470	4505	4610	4505	4435	4435	4435	4470	4505	4505	4540	4540
963	4505	4470	4470	4435	4470	4505	4540	4435	4540	4470	4470	4470	4540
976	4505	4435	4470	4540	4435	4470	4470	4470	4470	4540	4470	4540	4505
989	4535	4505	4575	4505	4505	4470	4505	4505	4435	4540	4505	4470	4470
1002	4470	4470	4505	4540	4470	4540	4470	4505	4540	4540	4505	4470	4470

No	4- rangée												
1015	4505	4540	4540	4505	4470	4400	4505	4470	4470	4505	4470	4470	4540
1028	4540	4505	4540	4540	4540	4505	4435	4435	4470	4505	4435	4505	4505
1041	4470	4505	4575	4400	4470	4505	4505	4470	4470	4505	4505	4540	4505
1054	4540	4435	4610	4540	4435	4435	4470	4505	4505	4470	4540	4540	4505
1067	4505	4435	4540	4470	4365	4434	4470	4505	4435	4470	4470	4505	4435
1080	4470	4540	4470	4435	4470	4470	4470	4470	4470	4505	4470	4540	4505
1093	4435	4470	4505	4470	4575	4470	4540	4470	4435	4470	4505	4505	4470
1106	4435	4575	4470	4470	4470	4435	4435	4505	4470	4435	4435	4540	4505
1119	4435	4435	4470	4400	4540	4575	4575	4505	4400	4470	4470	4505	4435
1132	4470	4435	4505	4435	4435	4435	4505	4505	4470	4540	4470	4575	4505
1145	4470	4470	4505	4505	4470	4470	4505	4575	4505	4470	4470	4540	4470
1158	4575	4505	4470	4540	4505	4505	4470	4505	4505	4435	4470	4470	4435
1171	4505	4435	4470	4435	4470	4435	4540	4505	4470	4400	4540	4470	4540
1184	4470	4505	4470	4470	4470	4505	4540	4470	4470	4540	4470	4505	4435
1197	4505	4505	4470	4470	4435	4505	4540	4505	4470	4505	4470	4575	4505
1210	4505	4505	4505	4540	4470	4435	4540	4470	4505	4540	4575	4505	4435
1223	4505	4470	4540	4505	4505	4505	4540	4505	4470	4505	4505	4435	4505
1236	4435	4505	4470	4470	4505	4470	4505	4505	4540	4505	4505	4505	4505
1249	4470	4505	4505	4505	4470	4470	4470	4540	4470	4505	4505	4470	4540
1262	4470	4540	4575	4470	4470	4470	4505	4470	4505	4470	4470	4540	4470
1275	4470	4505	4540	4435	4470	4470	4470	4505	4540	4540	4470	4505	4505
1288	4540	4470	4575	4400	4505	4540	4470	4505	4505	4470	4470	4540	4505
1301	4540	4540	4470	4470	4470	4505	4470	4505	4470	4470	4435	4435	4435
1314	4505	4435	4540	4540	4470	4435	4435	4470	4505	4470	4435	4470	4435
1327	4435	4470	4540	4435	4435	4540	4505	4470	4435	4540	4435	4470	4470
1340	4540	4435	4505	4540	4505	4470	4505	4540	4505	4470	4435	4470	4470

No	5- rangée												
1353	4435	4470	4505	4540	4505	4470	4470	4505	4540	4575	4435	4540	4470
1366	4470	4435	4470	4470	4470	4470	4505	4575	4505	4575	4505	4435	4505
1379	4435	4505	4505	4470	4540	4505	4470	4505	4540	4470	4470	4470	4505
1392	4505	4505	4540	4435	4470	4470	4540	4435	4435	4470	4470	4505	4505
1405	4435	4505	4470	4435	4470	4470	4540	4470	4470	4470	4505	4505	4505
1418	4540	4505	4505	4400	4435	4505	4470	4540	4470	4470	4435	4505	4470
1431	4435	4470	4505	4505	4435	4470	4540	4470	4505	4470	4540	4505	4505
1444	4470	4505	4505	4470	4470	4505	4540	4470	4470	4505	4435	4470	4435
1457	4540	4505	4435	4575	4470	4540	4435	4470	4470	4470	4470	4470	4505
1470	4505	4505	4540	4435	4505	4435	4470	4505	4540	4505	4540	4505	4470
1483	4505	4505	4435	4470	4470	4505	4435	4505	4505	4505	4505	4505	4435
1496	4470	4540	4470	4400	4400	4540	4470	4470	4540	4470	4470	4470	4435
1509	4470	4505	4470	4505	4505	4540	4470	4505	4470	4540	4470	4505	4505
1522	4435	4540	4470	4435	4505	4470	4470	4470	4435	4540	4435	4435	4540
1535	4540	4470	4470	4470	4505	4505	4470	4575	4505	4470	4540	4435	4505
1548	4505	4435	4435	4505	4540	4505	4400	4470	4505	4540	4505	4505	4470
1561	4435	4505	4470	4470	4435	4505	4505	4505	4540	4540	4505	4470	4470
1574	4505	4505	4435	4505	4540	4435	4470	4470	4505	4470	4505	4435	4505
1587	4505	4435	4505	4400	4505	4470	4505	4470	4470	4400	4505	4470	4540
1600	4470	4540	4505	1540	4505	4470	4505	4470	4540	4505	4505	4470	4470
1613	4470	4505	4505	4470	4435	4505	4435	4505	4470	4470	4505	4470	4540
1626	4470	4435	4470	4470	4505	4505	4470	4505	4505	4505	4435	4470	4505
1639	4540	4435	4470	4540	4435	4470	4505	4470	4470	4540	4505	4470	4435
1652	4540	4505	4470	4470	4505	4505	4400	4505	4505	4505	4505	4505	4505
1665	4470	4470	4505	4575	4505	4435	4400	4505	4470	4540	4470	4505	4505
1678	4400	4470	4435	4505	4435	4435	4540	4505	4505	4470	4470	4505	4575

No	6- rangée												
1691	4505	4505	4435	4505	4470	4540	4470	4505	4505	4540	4540	4435	4470
1704	4505	4505	4470	4505	4505	4470	4505	4435	4505	4470	4540	4470	4505
1717	4505	4540	4540	4435	4470	4505	4505	4470	4505	4505	4470	4505	4505
1730	4470	4470	4435	4400	4505	4505	4505	4505	4540	4505	4470	4470	4505
1743	4540	4505	4505	4540	4505	4505	4505	4505	4400	4505	4540	4505	4470
1756	4540	4505	4505	4505	4505	4435	4470	4505	4470	4505	4505	4470	4505
1769	4400	4470	4470	4505	4435	4505	4505	4470	4540	4505	4505	4505	4505
1782	4400	4505	4505	4505	4470	4470	4540	4400	4505	4505	4505	4505	4470
1795	4400	4470	4505	4505	4470	4505	4540	4470	4470	4505	4470	4470	4470
1808	4540	4540	4575	4505	4505	4505	4435	4470	4505	4540	4505	4470	4435
1821	4505	4470	4540	4435	4575	4435	4505	4435	4505	4540	4400	4470	4540
1834	4470	4505	4470	4505	4505	4470	4505	4470	4505	4505	4505	4470	4400
1847	4435	4400	4505	4470	4505	4435	4470	4505	4400	4470	4470	4470	4505
1860	4505	4470	4470	4540	4540	4470	4470	4540	4505	4505	4470	4470	4540
1873	4400	4470	4540	4470	4505	4505	4470	4470	4540	4470	4470	4470	4470
1886	4505	4505	4540	4540	4470	4505	4470	4505	4470	4470	4435	4470	4540
1899	4540	4435	4540	4470	4505	4470	4505	4470	4470	4505	4470	4470	4470
1912	4540	4470	4470	4505	4505	4470	4470	4435	4470	4505	4470	4505	4470
1925	4470	4575	4540	4505	4505	4470	4505	4470	4365	4540	4505	4470	4505
1938	4470	4470	4470	4470	4540	4505	4470	4540	4505	4435	4435	4540	4435
1951	4470	4505	4505	4470	4470	4470	4505	4470	4470	4470	4435	4470	4540
1964	4505	4435	4470	4470	4435	4505	4470	4400	4540	4470	4470	4505	4435
1977	4540	4470	4470	4505	4575	4470	4505	4470	4505	4540	4540	4435	4470
1990	4540	4435	4470	4505	4505	4575	4470	4540	4505	4505	4505	4470	4470
2003	4540	4540	4470	4470	4505	4470	4470	4505	4505	4505	4470	4470	4470
2016	4505	4435	4470	4505	4540	4505	4540	4435	4505	4540	4470	4470	4505

No	7- rangée												
2029	4470	4505	4470	4505	4505	4505	4575	4505	4435	4470	4435	4470	4505
2042	4435	4505	4505	4505	4505	4505	4470	4505	4505	4505	4435	4505	4435
2055	4435	4505	4470	4505	4435	4505	4505	4470	4470	4470	4505	4435	4470
2068	4365	4470	4505	4505	4435	4540	4470	4505	4505	4435	4505	4505	4505
2081	4470	4435	4435	4505	4540	4505	4505	4400	4505	4470	4540	4505	4470
2094	4505	4470	4470	4505	4470	4505	4505	4470	4470	4470	4575	4505	4470
2107	4435	4470	4540	4435	4505	4540	4540	4540	4435	4435	4470	4470	4470
2120	4540	4540	4470	4470	4540	4470	4435	4470	4505	4435	4400	4540	4540
2133	4470	4435	4470	4505	4470	4505	4505	4435	4505	4470	4470	4435	4505
2146	4435	4540	4470	4505	4505	4470	4470	4505	4470	4470	4540	4470	4435
2159	4470	4470	4505	4470	4505	4470	4500	4540	4540	4435	4505	4400	4470
2172	4505	4505	4470	4505	4470	4470	4435	4470	4470	4505	4435	4505	4470
2185	4470	4575	4505	4470	4435	4505	4435	4540	4505	4505	4435	4435	4505
2198	4505	4505	4575	4505	4505	4470	4470	4505	4470	4505	4470	4400	4540
2211	4505	4470	4470	4540	4470	4435	4470	4470	4400	4505	4540	4470	4540
2224	4540	4540	4505	4540	4540	4470	4540	4505	4505	4470	4470	4470	4505
2237	4470	4470	4435	4470	4470	4505	4505	4540	4505	4470	4505	4435	4505
2250	4505	4505	4505	4540	4470	4505	4470	4470	4505	4540	4435	4470	4470
2263	4505	4540	4575	4505	4540	4470	4470	4470	4575	4470	4540	4540	4505
2276	4540	4470	4435	4575	4540	4505	4470	4470	4470	4470	4505	4435	4505
2289	4575	4470	4505	4470	4470	4505	4470	4435	4470	4505	4505	4505	4470
2302	4470	4400	4470	4470	4470	4540	4435	4470	4505	4505	4470	4505	4470
2315	4435	4540	4505	4505	4435	4505	4540	4505	4505	4470	4435	4470	4505
2328	4505	4505	4505	4470	4505	4505	4540	4435	4505	4505	4470	4505	4540
2341	4435	4470	4470	4505	4505	4505	4540	4505	4540	4505	4470	4435	4505
2354	4505	4505	4540	4505	4470	4540	4435	4540	4470	4505	4435	4435	4505

No	8- rangée												
2367	4470	4540	4505	4505	4470	4470	4470	4540	4470	4540	4505	4540	4505
2380	4470	4470	4505	4505	4470	4540	4435	4470	4470	4540	4470	4470	4540
2393	4505	4435	4470	4470	4505	4435	4505	4505	4575	4540	4470	4505	4540
2406	4575	4470	4435	4505	4505	4505	4435	4505	4505	4505	4470	4505	4400
2419	4470	4470	4470	4505	4435	4470	4435	4505	4435	4470	4470	4505	4540
2432	4470	4470	4540	4435	4540	4540	4505	4505	4505	4505	4470	4470	4540
2445	4505	4540	4435	4540	4470	4505	4470	4435	4540	4470	4540	4505	4435
2458	4435	4470	4540	4540	4505	4540	4435	4540	4575	4540	4575	4540	4540
2471	4505	4505	4470	4470	4505	4435	4505	4470	4505	4470	4505	4505	4470
2484	4470	4505	4505	4505	4505	4470	4470	4505	4505	4505	4435	4505	4470
2497	4505	4540	4470	4505	4470	4540	4505	4540	4470	4470	4505	4505	4505
2510	4470	4505	4470	4470	4505	4505	4540	4505	4540	4505	4505	4505	4470
2523	4575	4505	4435	4470	4435	4435	4470	4470	4435	4470	4470	4435	4540
2536	4505	4505	4505	4540	4470	4435	4470	4435	4400	4470	4435	4470	4505
2549	4435	4435	4470	4540	4470	4470	4435	4470	4505	4470	4435	4435	4470
2562	4470	4435	4505	4435	4505	4470	4470	4470	4470	4505	4505	4470	4540
2575	4505	4470	4435	4470	4540	4435	4505	4470	4435	4470	4540	4470	4540
2588	4470	4540	4505	4470	4505	4505	4470	4470	4435	4470	4470	4435	4435
2601	4505	4505	4470	4470	4435	4470	4575	4470	4505	4470	4470	4505	4505
2614	4435	4470	4435	4505	4505	4435	4470	4505	4540	4540	4505	4400	4470
2627	4470	4470	4505	4505	4505	4400	4435	4540	4435	4470	4540	4435	4505
2640	4470	4505	4505	4540	4505	4470	4505	4505	4505	4505	4470	4435	4505
2653	4505	4435	4540	4540	4505	4540	4505	4505	4505	4575	4505	4470	4470
2666	4470	4505	4470	4540	4540	4505	4470	4470	4540	4470	4435	4470	4540
2679	4470	4435	4505	4540	4470	4470	4470	4505	4505	4540	4540	4540	4435
2692	4505	4505	4505	4505	4505	4435	4435	4505	4470	4470	4505	4505	4540

No	9- rangée												
2705	4470	4505	4435	4470	4540	4505	4435	4505	4470	4470	4505	4470	4470
2718	4505	4505	4470	4505	4470	4505	4575	4505	4470	4505	4435	4470	4470
2731	4470	4435	4470	4540	4470	4470	4505	4505	4540	4540	4435	4540	4470
2744	4505	4505	4435	4470	4470	4505	4575	4435	4470	4505	4505	4505	4435
2757	4400	4435	4435	4505	4505	4470	4470	4505	4470	4505	4505	4435	4435
2770	4435	4470	4470	4505	4505	4470	4505	4505	4470	4505	4505	4435	4540
2783	4435	4540	4540	4505	4505	4505	4470	4540	4575	4505	4505	4505	4435
2796	4435	4540	4505	4540	4505	4435	4575	4470	4540	4470	4505	4540	4470
2809	4470	4505	4505	4540	4470	4505	4540	4540	4470	4505	4470	4400	4505
2822	4400	4470	4435	4435	4505	4470	4540	4400	4435	4470	4435	4470	4470
2835	4505	4505	4505	4505	4470	4470	4505	4505	4470	4505	4470	4470	4435
2848	4505	4470	4470	4505	4505	4505	4470	4470	4505	4435	4470	4505	4505
2861	4505	4435	4470	4470	4470	4470	4400	4540	4540	4505	4505	4540	4470
2874	4400	4470	4470	4400	4505	4470	4505	4470	4540	4505	4435	4540	4505
2887	4470	4505	4470	4470	4540	4470	4470	4505	4470	4505	4505	4470	4505
2900	4470	4505	4540	4505	4470	4505	4540	4505	4470	4505	4470	4470	4470
2913	4505	4470	4470	4435	4505	4400	4540	4540	4540	4435	4435	4505	4400
2926	4470	4505	4505	4470	4505	4540	4505	4470	4540	4505	4470	4470	4540
2939	4505	4470	4470	4505	4470	4505	4505	4505	4505	4470	4505	4470	4470
2952	4540	4505	4505	4470	4540	4540	4470	4470	4470	4435	4505	4400	4470
2965	4470	4470	4505	4435	4505	4505	4575	4435	4540	4470	4505	4435	4435
2978	4505	4470	4505	4470	4470	4470	4505	4470	4540	4470	4505	4470	4435
2991	4470	4470	4470	4470	4505	4435	4540	4470	4470	4505	4470	4505	4540
3004	4400	4540	4540	4505	4505	4470	4505	4505	4470	4505	4505	4435	4470
3017	4505	4470	4435	4540	4435	4505	4575	4435	4435	4505	4470	4470	4505
3030	4540	4470	4435	4435	4540	4470	4470	4470	4540	4470	4470	4505	4400

No	10- rangée												
3043	4470	4435	4505	4470	4470	4505	4505	4435	4540	4505	4435	4470	4470
3056	4540	4435	4470	4505	4470	4470	4505	4540	4505	4470	4505	4435	4435
3069	4505	4505	4470	4505	4470	4470	4540	4470	4470	4505	4435	4505	4540
3082	4505	4435	4470	4435	4470	4470	4505	4435	4505	4470	4505	4505	4435
3095	4470	4470	4540	4540	4470	4505	4505	4540	4505	4540	4470	4505	4435
3108	4540	4470	4540	4505	4540	4505	4505	4470	4400	4470	4470	4470	4505
3121	4470	4505	4470	4470	4540	4435	4505	4470	4540	4435	4540	4505	4470
3134	4470	4540	4470	4505	4435	4470	4505	4470	4470	4470	4435	4470	4435
3147	4505	4505	4505	4470	4435	4505	4505	4470	4505	4435	4505	4505	4470
3160	4505	4470	4505	4505	4505	4470	4435	4505	4505	4505	4470	4470	4470
3173	4435	4505	4435	4470	4575	4540	4435	4505	4505	4470	4435	4505	4505
3186	4400	4505	4505	4470	4505	4470	4435	4505	4470	4540	4505	4435	4470
3199	4575	4470	4470	4505	4505	4505	4505	4505	4505	4505	4505	4470	4470
3212	4470	4435	4470	4470	4505	4470	4435	4505	4540	4505	4505	4505	4470
3225	4435	4470	4540	4470	4540	4470	4470	4470	4470	4470	4540	4470	4540
3238	4470	4470	4470	4470	4470	4505	4505	4575	4540	4540	4400	4470	4505
3251	4505	4470	4505	4470	4470	4470	4470	4470	4575	4470	4505	4470	4540
3264	4470	4505	4470	4435	4470	4470	4505	4400	4435	4435	4435	4505	4505
3277	4505	4470	4435	4435	4470	4470	4505	4505	4505	4540	4470	4470	4470
3290	4435	4505	4505	4435	4435	4505	4505	4505	4505	4470	4540	4435	4540
3303	4470	4470	4470	4470	4470	4540	4470	4505	4505	4470	4505	4470	4505
3316	4470	4505	4505	4505	4400	4470	4435	4505	4435	4435	4505	4470	4470
3329	4470	4470	4540	4540	4470	4505	4505	4435	4470	4505	4470	4505	4505
3342	4470	4505	4470	4470	4435	4505	4505	4400	4505	4505	4540	4470	4470
3355	4470	4575	4470	4470	4470	4505	4470	4470	4505	4435	4435	4470	4540
3368	4470	4540	4505	4435	4470	4470	4540	4470	4435	4505	4505	4435	4505

No	11- rangée												
3381	4435	4505	4435	4470	4470	4435	4470	4540	4540	4470	4435	4505	4575
3394	4540	4505	4505	4540	4470	4435	4505	4435	4470	4505	4470	4470	4505
3407	4505	4505	4470	4470	4470	4505	4505	4470	4435	4435	4540	4435	4505
3420	4470	4505	4505	4470	4505	4505	4435	4505	4470	4470	4505	4435	4470
3433	4470	4470	4540	4540	4540	4470	4435	4470	4505	4505	4470	4435	4540
3446	4505	4505	4505	4470	4540	4505	4505	4400	4505	4505	4435	4450	4470
3459	4470	4470	4540	4470	4505	4505	4540	4505	4435	4470	4470	4470	4505
3472	4470	4470	4505	4470	4470	4505	4470	4470	4505	4540	4435	4505	4400
3485	4470	4505	4505	4540	4470	4470	4505	4435	4435	4470	4470	4540	4470
3498	4505	4540	4540	4505	4505	4575	4470	4505	4540	4470	4505	4435	4540
3511	4505	4470	4470	4540	4505	4505	4435	4540	4470	4505	4540	4540	4470
3524	4505	4505	4470	4505	4540	4470	4470	4505	4505	4505	4540	4505	4505
3537	4505	4505	4505	4470	4470	4470	4435	4435	4470	4540	4505	4470	4540
3550	4470	4435	4470	4470	4470	4540	4540	4505	4505	4505	4505	4505	4470
3563	4470	4435	4505	4505	4505	4505	4470	4435	4470	4470	4505	4505	4470
3576	4470	4505	4435	4505	4505	4470	4505	4505	4470	4470	4505	4470	4505
3589	4505	4435	4505	4470	4470	4470	4575	4470	4470	4575	4470	4470	4470
3602	4435	4470	4470	4540	4470	4505	4470	4400	4505	4470	4470	4470	4435
3615	4470	4505	4540	4470	4505	4470	4470	4470	4470	4470	4435	4470	4505
3628	4505	4505	4470	4575	4400	4470	4435	4540	4505	4470	4540	4400	4505
3641	4505	4470	4540	4435	4505	4470	4470	4470	4470	4505	4505	4470	4470
3654	4470	4435	4435	4505	4470	4470	4505	4470	4505	4505	4540	4505	4505
3667	4470	4470	4575	4435	4470	4435	4470	4400	4470	4470	4470	4470	4505
3680	4540	4470	4575	4540	4505	4470	4540	4435	4470	4435	4470	4505	4505
3693	4470	4505	4505	4505	4470	4470	4470	4505	4575	4435	4505	4470	4470
3706	4470	4505	4470	4505	4470	4470	4435	4505	4435	4505	4505	4470	4505

No

12- rangée

3719	4540	4470	4505	4470	4435	4470	4470	4505	4505	4470	4505	4505	4470
3732	4470	4435	4505	4470	4435	4470	4435	4540	4540	4470	4470	4470	4540
3745	4435	4505	4470	4540	4505	4435	4470	4540	4470	4540	4435	4470	4470
3758	4470	4470	4470	4540	4470	4505	4435	4470	4575	4540	4470	4540	4470
3771	4470	4470	4540	4470	4470	4505	4470	4505	4470	4505	4470	4540	4575
3784	4470	4540	4470	4505	4505	4505	4435	4505	4505	4505	4470	4505	4470
3797	4470	4540	4505	4505	4435	4505	4575	4470	4435	4505	4470	4505	4435
3810	4540	4540	4505	4575	4470	4505	4505	4470	4470	4505	4505	4470	4505
3823	4505	4505	4470	4505	4435	4540	4470	4505	4505	4505	4505	4540	4435
3836	4505	4540	4505	4505	4470	4470	4470	4540	4505	4470	4540	4400	4505
3849	4470	4540	4505	4540	4540	4470	4470	4540	4505	4435	4505	4540	4505
3862	4505	4435	4470	4470	4435	4435	4540	4505	4505	4470	4505	4435	4505
3875	4470	4505	4505	4435	4470	4505	4470	4470	4470	4470	4505	4470	4435
3888	4505	4470	4505	4505	4470	4505	4505	4505	4470	4540	4505	4435	4400
3901	4435	4505	4540	4505	4470	4470	4540	4470	4470	4540	4505	4540	4505
3914	4505	4505	4435	4505	4470	4540	4540	4435	4505	4505	4505	4540	4540
3927	4505	4435	4540	4505	4505	4435	4540	4505	4435	4505	4505	4505	4505
3940	4470	4540	4435	4540	4435	4505	4505	4435	4540	4505	4435	4540	4505
3953	4540	4505	4505	4505	4505	4505	4540	4435	4505	4435	4540	4505	4540
3966	4435	4505	4505	4435	4505	4505	4435	4435	4435	4435	4505	4505	4505
3979	4470	4505	4435	4505	4505	4505	4505	4435	4435	4435	4435	4435	4435
3992	4435	4435	4435	4435	4435	4435	4435	4435	4435				

Le poids moyen des sacs estimé par le gérant se situe entre 4495 g et 4515 g et le pourcentage de sacs pourris entre 23% et 37%. Fort surpris de la valeur de ces estimations, il communique avec le producteur-grossiste.

En effet, le contrat ne semble pas respecté par le producteur. Selon l'estimation faite par le gérant, la majorité des sacs pèsent autour de 4505 g au lieu de 4540 g et quelques-uns n'atteignent pas la limite de 4400 g fixée dans le contrat. De plus, le pourcentage de sacs pourris tourne autour de 30% ce qui dépasse largement le 10% maximum mentionné dans le contrat.

Le producteur, un peu sceptique devant ces résultats, avant de reprendre tous les sacs livrés et d'en contrôler le poids et la qualité un à un, opération longue et coûteuse, propose comme compromis au gérant de prendre un deuxième échantillon. Il s'engage cependant à reprendre tous les sacs à condition que la méthode d'échantillonnage soit suffisamment efficace, c'est-à-dire qu'il ait peu de chance (au plus 0,25%) d'être obligé de reprendre sa marchandise alors qu'elle rencontre les normes.

Le gérant, de son côté, conscient qu'un échantillon peut donner une valeur proche de 4540 g ou de 10% sans que sa première estimation soit réellement fausse, ne peut accepter une livraison de sacs dont le poids ne serait pas juste, ou le pourcentage de sacs pourris trop élevé, en tous cas, au plus dans 0,25% lui aussi.

Calculer la taille de l'échantillon qui satisfait les conditions du producteur et du gérant.

Construire cet échantillon.

1. *Base de sondage*

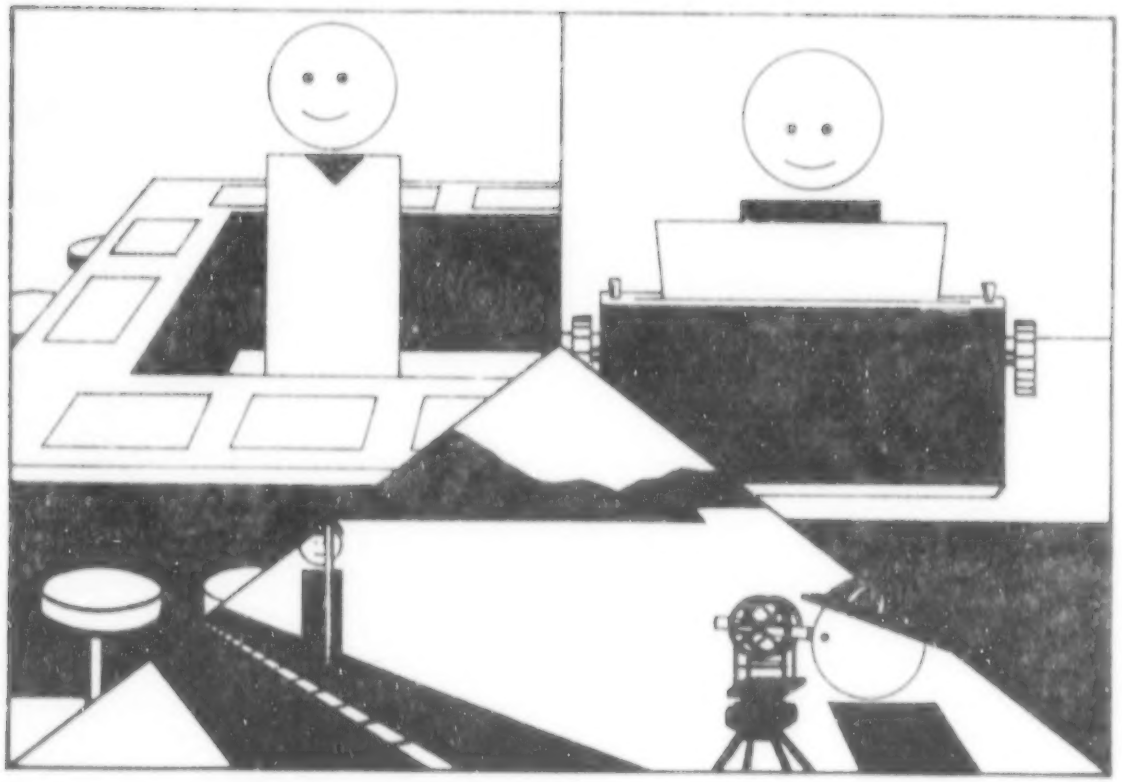
2. *Lecture de la table de nombres au hasard (1ère rangée, etc.)*

3. *L'échantillon*

Q 4. *Le producteur doit-il reprendre ses sacs?*

Expliquer.

Étude du revenu des étudiants(es) en statistique au cours des dernières vacances estivales



APPLICATION IV

NOM(S) : _____

_____

_____

OBJECTIF GÉNÉRAL

Acquérir les concepts statistiques et développer les habiletés nécessaires à l'analyse des données.

MATÉRIEL

Un questionnaire.

CONTEXTE DE L'APPLICATION

Nous voulons étudier les activités des étudiants(es) en statistique au cours des dernières vacances estivales. Cette étude fait appel à la statistique descriptive, aux probabilités et à la théorie d'échantillonnage. Pour cette étude on suivra la méthodologie employée à l'application II.

CHEMINEMENT

L'enquête précédente a répondu à des questions concernant les gens de la classe mais tu as pu te demander au cours de ton travail si les résultats obtenus étaient applicables à l'extérieur de la classe, c'est-à-dire à tout le C.E.G.E.P., à la région, au Québec ou ailleurs.

Supposons que nous voulons connaître les activités estivales des étudiants(es) du C.E.G.E.P.; examinons quelques hypothèses de travail pour effectuer cette recherche:

1. Faire parvenir un questionnaire à tous (toutes) les étudiants(es) du C.E.G.E.P.; cette solution paraît difficile vu le très grand nombre de personnes à rejoindre et qu'en plus, plusieurs pourraient ne pas répondre (le problème des non-répondants est complexe étant donné qu'on ne sait pas s'ils constituent un groupe à part ou s'ils sont à peu près comme ceux qui ont répondu).

2. Faire parvenir un questionnaire à un nombre plus restreint de personnes choisies selon les pourcentages de masculin, féminin, professionnel, général... que l'on retrouve au C.E.G.E.P.. Cette solution quoique moins onéreuse que la précédente exige de connaître déjà passablement la population étudiante du C.E.G.E.P. en plus du travail requis pour choisir ces personnes pour ensuite leur soumettre le questionnaire. (Cette méthode s'appelle échantillonnage par stratification).

3. Faire parvenir un questionnaire à un échantillon d'étudiants(es) choisis(es) au hasard sur la liste des étudiants(es) du C.E.G.E.P.. (Cette méthode s'appelle échantillonnage aléatoire simple.) Elle paraît plus simple mais elle exige quand même de rejoindre différents(es) étudiants(es) de la collectivité collégiale; en effet, les étudiants(es) de la classe ou ceux(celles) que l'on peut rencontrer à la porte de la cafétéria ne constituent pas un échantillon aléatoire simple de la population du C.E.G.E.P.

Dans le contexte actuel d'un cours d'initiation en statistique, nous allons restreindre la population à l'ensemble des étudiants(es) qui suivent des cours de statistique durant cette session-ci; l'échantillon sera choisi parmi ces étudiants(es) et par conséquent, nos résultats seront applicables seulement à l'ensemble des étudiants(es) qui suivent des cours de statistique et non à la population du C.E.G.E.P. en général.

Problématique

Le Service des prêts et bourses du ministère de l'Éducation octroie des prêts et/ou des bourses aux étudiants(es) en supposant que l'ensemble des étudiants(es) ont travaillé et gagné le salaire minimum.

Ces suppositions sont-elles réalistes, du moins pour les étudiants(es) en statistique de notre collège?

Nous pouvons traduire cette interrogation sous la forme de quatre questions:

Q 1. *Quel a été le revenu total moyen des étudiants(es) de statistique au cours des dernières vacances estivales?*

Q 2. *Le revenu hebdomadaire d'été des étudiants(es) de statistique est-il inférieur au revenu hebdomadaire minimum de base estimé par le Service des prêts et bourses?*

Q 3. *La majorité des étudiants(es) de statistique ont-il travaillé au cours des dernières vacances estivales?*

Q 4. *Quelle proportion d'étudiants(es) de statistique ont travaillé au cours des dernières vacances estivales?*

Méthodologie

Afin de répondre aux quatre questions, il faut choisir au hasard un échantillon d'étudiants(es) parmi l'ensemble des étudiants(es) qui suivent un cours de statistique durant cette session-ci.

Le nombre de personnes qui feront partie de l'échantillon sera déterminé par les conditions spécifiques à chaque question.

Pour recueillir l'information pertinente, tu devras utiliser les réponses fournies par les étudiants(es) au début de la session.

Cueillette des données

Puisque tu vas utiliser les réponses fournies par les étudiants(es) au début de la session, la cueillette des données est déjà faite.

Tableau d'observations et codification

Lorsque les personnes qui vont faire partie de l'échantillon seront identifiées, tu pourras remplir un tableau d'observations. La codification a déjà été faite.

Analyse statistique des données

Il faut essayer de répondre aux questions l'une après l'autre.

Q 1. Quel a été le revenu total moyen des étudiants(es) de statistique au cours des dernières vacances estivales?

Nous estimerons ce revenu total à l'aide d'un échantillon.

Comment choisir les éléments de l'échantillon?

CONSULTER LE MANUEL D'INSTRUCTIONS À CE SUJET.

On veut connaître le revenu total moyen avec une précision de plus ou moins 200 $ et un niveau de confiance de 95% en nos résultats. Combien d'éléments devrons-nous choisir pour construire l'échantillon?

CONSULTER LE MANUEL D'INSTRUCTIONS À CE SUJET.

précision = 200 $ = $z_c \, \sigma_{\bar{x}}$

Q 2. Le revenu hebdomadaire d'été des étudiants(es) de statistique à cette session-ci est-il inférieur au revenu hebdomadaire minimum de base estimé par le Service des prêts et bourses?

Le gouvernement présume que les étudiants(es) ont eu un revenu hebdomadaire minimum de ..... $; nous pouvons essayer de vérifier si ce revenu a été atteint par les étudiants(es) de statistique lors des dernières vacances estivales.

Le problème de vérification de la validité d'une hypothèse vient bien sûr du fait que la moyenne d'un échantillon n'est pas nécessairement égale à la moyenne de la population dont il est tiré.

Comment tester une hypothèse?

CONSULTER LE MANUEL D'INSTRUCTIONS À CE SUJET.

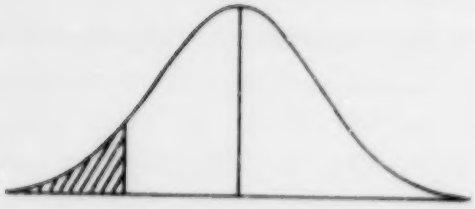
Donc les hypothèses sont:

H_0: .....

H_1: .....

Comme tu l'as vu dans le manuel d'instructions, il faut maintenant fixer un seuil de signification; prenons ici un seuil de 0,05.

Établissons la règle de décision; il s'agit d'un test unilatéral.



Utiliser le même échantillon qu'à la question 1.

Cet échantillon permet-il d'accepter l'hypothèse H_0 ou l'hypothèse H_1?

Q 8. La majorité des étudiants(es) de statistique à cette session-ci ont-ils travaillé au cours des dernières vacances estivales?

Comment tester une hypothèse?

CONSULTER LA MANUEL D'INSTRUCTIONS À CE SUJET.

Vérifier cette hypothèse.

Q 4. Quelle proportion d'étudiants(es) de statistique à cette session-ci ont travaillé au cours des dernières vacances estivales?

Pour des raisons pratiques, nous utiliserons le même échantillon qu'à la question 1. Utiliser un niveau de confiance de 99%; remarquer que la précision de l'estimation résultera de la nature et de la taille de l'échantillon.

Comment estimer un paramètre d'une population?

CONSULTER LE MANUEL D'INSTRUCTIONS À CE SUJET.

Pour estimer la proportion de l'ensemble des étudiants(es) en statistique qui ont travaillé au cours des dernières vacances estivales, nous devrons utiliser:

$$p \pm z_c \, \sigma_p$$

Faire cette estimation.

En suivant la méthodologie déjà vue, faire un rapport sur le sujet.

3. Travaux pratiques

T1. Sructure temporelle

(Étude de la consommation électrique)

Dans l'industrie, les commerces ou les maisons d'enseignement, l'Hydro-Québec ne calcule pas toujours la comsommation électrique de la même façon que dans les résidences privées. Le mode de calcul est le suivant: on fixe un taux régulier de consommation avec une certaine marge de fluctuation, si cette marge n'est pas dépassée même si le taux est dépassé, il n'y a pas de surcharge; mais si cette marge est dépassée, il faut tout simplement majorer le taux ou payer une surcharge assez dispendieuse.

À l'hiver 1981, l'Université du Québec à Rimouski constatait qu'à certaines heures de la journée, sa consommation montait un peu trop et voulait corriger cette situation.

D'abord l'Hydro-Québec a voulu vérifier la situation exacte durant une semaine, en raccordant un compteur à l'entrée électrique de l'université, cela du 81-02-01 au 81-02-08.

Les résultats de cette lecture, tu les retrouves au tableau 1.1 page T 1.5

Placer ces résultats sur le graphique à la page T 1.7 (en identifiant chaque journée par une couleur différente), pour avoir une meilleure idée de la situation.

Indiquer à quelle(s) heure(s) tu remarques certaines pointes (pics).

Indiquer les pointes à éliminer.

Après cette étude, l'Hydro-Québec a conseillé à l'Université de fermer ses chauffe-eau de 10 h 30 à 11 h 30 et de 14 h 30 à 15 h 30, pour essayer de minimiser un risque d'augmentation de coût.

Le tableau 1.2 donne les lectures du compteur pour la semaine du 81-02-16 au 81-02-21 où on a fermé les chauffe-eau.

Placer ces résultats sur le graphique à la page T 1.8 (en identifiant chaque journée par une couleur différente).

Q 1. La correction apportée à la situation te semble-t-elle suffisante?

.....

Justifier cette réponse en indiquant l'endroit et la valeur des nouvelles pointes.

Q 2. Si l'hydro-Québec permet un jeu de 4% de la consommation, penses-tu que le taux de 800kW à l'heure risquerait d'être dépassé dans la situation corrigée?

.....

Justifier cette réponse en comparant les nouvelles pointes au maximum permis.

Tableau 1.1 La consommation électrique de l'Université du Québec à Rimouski du 81-02-01 au 81-02-08 (En kilowatts).

Journée / Heure	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi
1	390 kW	380 kW	290 kW	230 kW	240 kW	220 kW
2	-	-	-	-	-	-
3	300	260	250	320	260	240
4	-	-	-	-	-	-
5	300	280	270	340	280	250
6	-	-	-	-	-	-
7	290	320	270	280	280	200
8	400	420	500	430	220	220
9	560	580	560	550	360	220
10	560	540	560	540	400	240
11	600	660	560	520	360	220
12	500	520	500	520	400	240
13	520	580	540	620	440	240
14	540	600	600	620	360	260
15	540	680	640	620	280	260
16	500	580	580	590	280	240
17	420	460	510	490	280	250
18	360	460	530	600	270	250
19	440	460	550	620	260	260
20	400	520	460	470	260	260
21	340	460	420	400	260	260
22	-	-	-	-	-	-
23	320	320	380	340	260	250
24	-	-	-	-	-	-

Tableau 1.2 La consommation électrique de l'Université du Québec à Rimouski du 81-02-16 au 81-02-21 (après fermeture des chauffe-eau de 10 h 30 à 11 h 30 et de 14 h 30 à 15 h 30).

Journée / Heure	Lundi	Mardi	Mercredi	Jeudi	Vendredi
7	370 kW	320 kW	300 kW	280 kW	390 kW
8	570	430	400	430	430
9	620	580	590	560	500
10	600	570	550	580	490
11	570	560	600	580	520
12	580	570	500	540	480
13	530	530	520	540	460
14	560	570	560	570	520
15	570	560	560	540	490
16	580	590	570	590	500
17	520	500	540	560	480
18	540	540	520	540	-
19	520	500	480	530	-
20	500	460	440	480	-
21	480	440	460	460	-
22	460	360	490	440	-

ÉTUDE D'ÉCONOMIE D'ÉNERGIE

Bâtiments existants

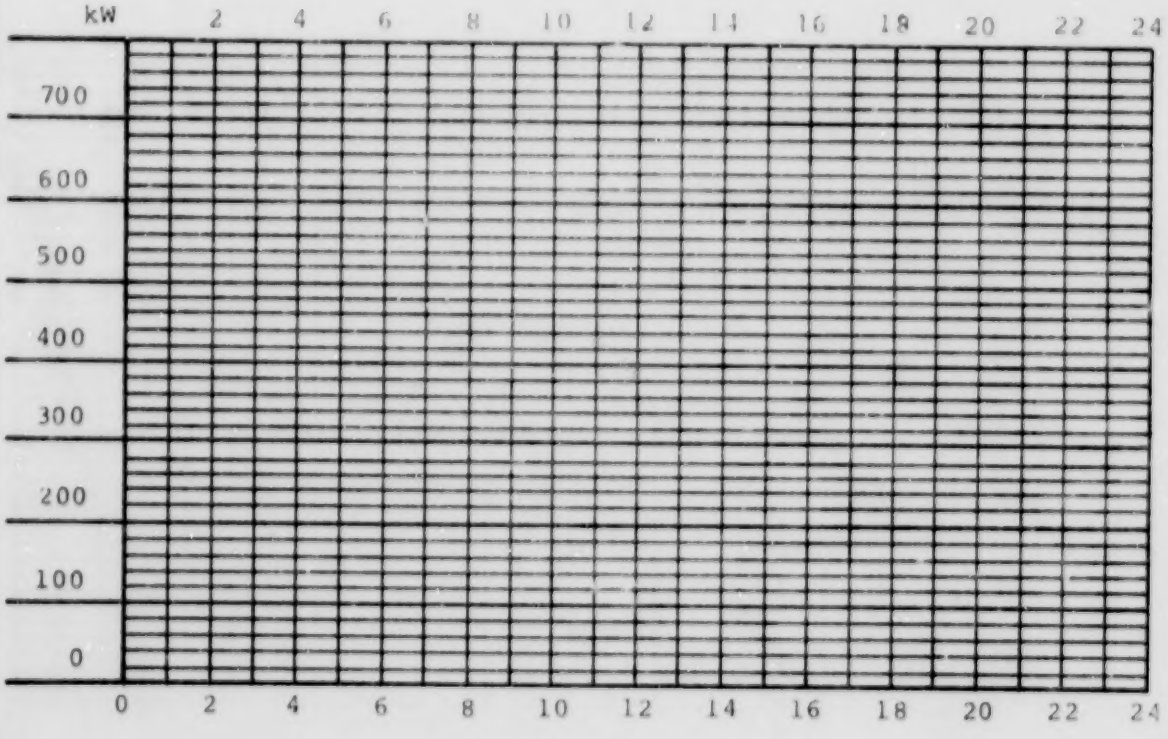
PROFIL D'UNE CHARGE ÉLECTRIQUE

Période du relevé 81-02-01 au 81-02-08

Légende (ou couleur)

Charges mesurées _____

- Lundi
- Mardi
- Mercredi
- Jeudi
- Vendredi
- Samedi
- Dimanche



Heure du jour

Date du relevé _____

Nom du releveur _____

ÉTUDE D'ÉCONOMIE D'ÉNERGIE

Bâtiments existants

PROFIL D'UNE CHARGE ÉLECTRIQUE

Période du relevé 81-02-16 au 81-02-21

Légende (ou couleur)

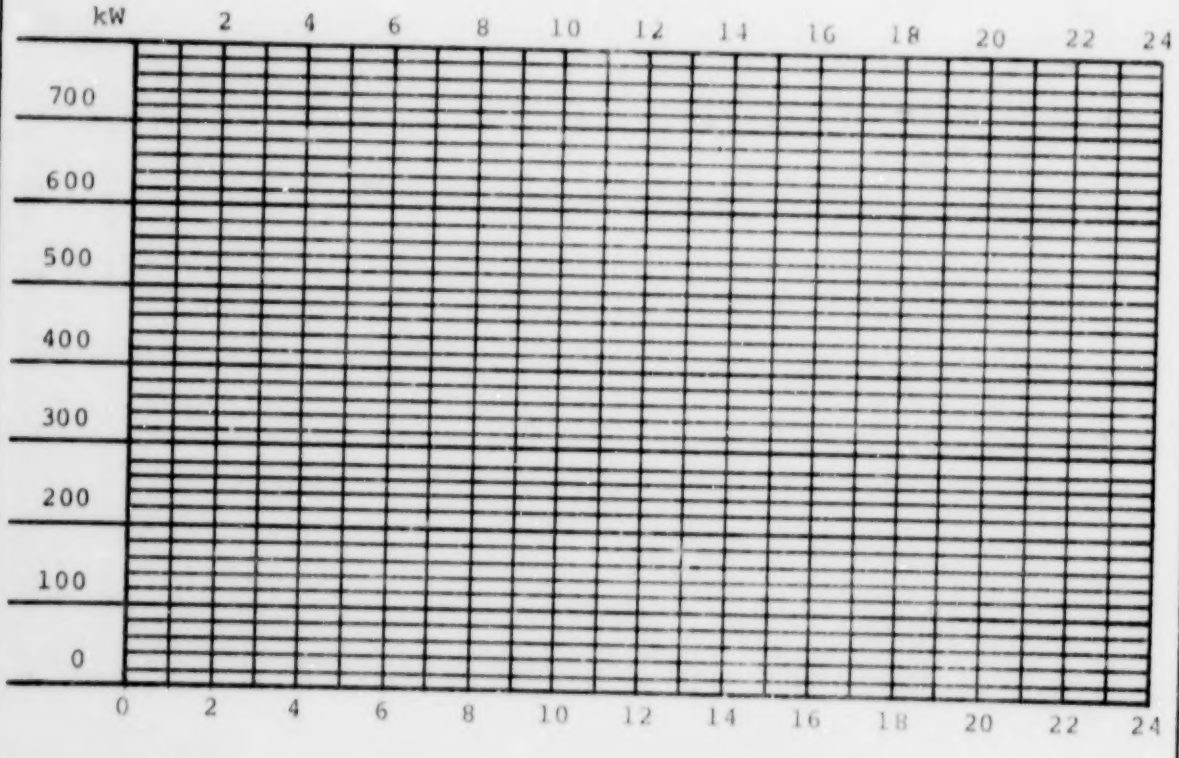
Charges mesurées _____

_____

_____

_____

Lundi
Mardi
Mercredi
Jeudi
Vendredi
Samedi
Dimanche



Heure du jour

Date du relevé _____

Nom du releveur _____

T2. Synthèse

(Étude d'une échelle de salaire)

Lors du renouvellement de leur convention collective, les employés d'hôpitaux veulent, dans leur demande salariale, garder une certaine qualité de vie en s'assurant une augmentation minimum, égale à l'augmentation du coût de la vie, soit 10%. En plus, ils veulent que cette augmentation ne grandisse pas l'écart, qui existe déjà entre les différents taux horaires, mais que cet écart puisse être diminué.

Les patrons, dans leurs offres, leur proposent une alternative, qui doit revenir au même coût:

Une augmentation de 11% à tous les employés ou pour se rapprocher de la demande syndicale de ne pas augmenter les écarts, un montant fixe et égal à tous.

Pour comparer la demande avec les offres, il faut prendre des exemples d'hôpitaux comme ceux indiqués dans les tableaux 2.8 et 2.9.

De plus, pour vérifier si la demande dépasse de beaucoup la première offre des patrons et aussi calculer le montant fixe d'augmentation de la deuxième offre, il faut évaluer la tendance centrale et la dispersion des taux horaires.

Travail sur le sujet

Calculer la moyenne et l'écart moyen des taux horaires actuels des employé(es) de chaque hôpital. (Inscrire les résultats dans le tableau suivant.)

Tableau 2.1 La moyenne et l'écart moyen des taux horaires actuels des employés(es) de chaque hôpital

	Hôpital I	Hôpital II
Moyenne		
Écart moyen		

Calculer la moyenne et l'écart moyen des taux horaires qu'obtiendraient les employé(es) de chaque hôpital s'ils acceptent la première offre patronale, qui se traduit par l'équation mathématique suivante:

Taux offert = 1.11 (taux actuel)

	Hôpital I	Hôpital II
Moyenne		
Écart moyen		

Calculer le montant total de l'augmentation pour les deux hôpitaux:

Montant total d'augmentation: .....
(première offre)

Calculer le montant fixe et uniforme que chaque employé(e) recevrait s'il(elle) accepte la deuxième offre patronale.

Montant uniforme d'augmentation à chaque employé(e):
.....

Calculer la moyenne et l'écart moyen des taux horaires selon la deuxième offre patronale.

Tableau 2.3 La moyenne et l'écart moyen des taux horaires des employés(es) de chaque hôpital selon la deuxième offre patronale

	Hôpital I	Hôpital II
Moyenne		
Écart moyen		

La demande syndicale peut se traduire par l'équation mathématique suivante:

Taux horaire demandé = 9/10 (taux actuel) + 2,00

Calculer la moyenne et l'écart moyen des taux horaires des employés(es) selon la demande syndicale.

Tableau 2.4 La moyenne et l'écart moyen des taux horaires des employés(es) de chaque hôpital selon la demande syndicale

	Hôpital I	Hôpital II
Moyenne		
Écart moyen		

Calculer le montant total d'augmentation de la demande syndicale pour les deux hôpitaux.

Montant total d'augmentation: .....
(demande syndicale)

Q 1. Du combien la demande syndicale dépasse-t-elle l'offre patronale?
.....

Calculer l'augmentation réelle dans les trois cas pour le salarié le mieux payé et souligner la situation la plus avantageuse.

Tableau 2.5 L'augmentation réelle du salarié le mieux payé de chaque hôpital

	Hôpital I	Hôpital II
1ière offre		
2e offre		
demande		

Calculer l'augmentation réelle dans les trois cas pour le salarié le mieux payé et souligner la situation la plus avantageuse.

Tableau 2.6 L'augmentation réelle du salarié le moins bien payé de chaque hôpital

	Hôpital I	Hôpital II
1ière offre		
2e offre		
demande		

Calculer l'augmentation réelle dans les trois cas pour le salarié médian et souligner la situation la plus avantageuse.

Tableau 2.7 L'augmentation réelle du salarié médian de chaque hôpital

	Hôpital I	Hôpital II
1ière offre		
2e offre		
demande		

Interpréter les résultats selon l'optique de l'employé(e) et de l'employeur.

TABLEAU 2.8

L'échelle de salaires des 75 employés de l'hôpital I

(On indique par un numéro chaque employé classé alphabétiquement.)

No	Taux horaire ($)	No	Taux horaire ($)	No	Taux horaire ($)
1	5,50	26	8,00	51	7,00
2	6,25	27	7,00	52	5,50
3	6,50	28	5,50	53	5,50
4	9,00	29	5,00	54	6,50
5	5,50	30	6,50	55	8,00
6	7,00	31	8,50	56	8,25
7	8,50	32	5,50	57	6,25
8	5,00	33	7,50	58	5,50
9	9,00	34	9,00	59	7,50
10	6,50	35	10,00	60	8,50
11	5,50	36	6,50	61	5,50
12	7,50	37	5,50	62	8,00
13	10,00	38	8,00	63	9,00
14	5,50	39	9,00	64	8,25
15	5,50	40	6,25	65	6,50
16	6,50	41	7,50	66	5,00
17	7,50	42	9,00	67	5,50
18	9,00	43	8,50	68	7,50
19	6,25	44	10,00	69	9,00
20	5,50	45	7,00	70	8,50
21	7,00	46	5,50	71	5,50
22	8,00	47	9,00	72	7,00
23	9,00	48	8,25	73	8,00
24	10,00	49	8,00	74	8,50
25	7,00	50	7,00	75	7,50

TABLEAU 2.9

L'échelle de salaires des 125 employés de l'hôpital II

No	Taux horaire ($)	No	Taux horaire ($)	No	Taux horaire ($)	No	Taux horaire ($)
1	7,75	33	8,00	65	7,00	97	5,00
2	8,00	34	7,50	66	6,00	98	5,75
3	8,25	35	6,00	67	8,50	99	6,50
4	8,50	36	8,25	68	8,00	100	9,25
5	7,00	37	8,75	69	8,25	102	6,00
6	8,75	38	8,00	70	7,50	102	9,50
7	7,75	39	8,25	71	6,50	103	5,50
8	8,25	40	7,75	72	9,25	104	7,00
9	8,00	41	8,50	73	6,00	105	6,50
10	8,00	42	9,00	74	9,50	106	8,00
11	8,50	43	8,75	75	6,50	107	8,25
12	7,50	44	8,00	76	5,75	108	7,50
13	9,00	45	8,00	77	5,50	109	8,50
14	8,75	46	8,25	78	9,25	110	8,00
15	8,00	47	7,75	79	6,50	111	8,00
16	8,25	48	8,50	80	6,00	112	8,25
17	7,75	49	7,50	81	5,00	113	7,75
18	7,00	50	7,00	82	5,75	114	8,75
19	8,00	51	8,75	83	9,25	115	7,50
20	10,00	52	8,25	84	6,50	116	9,00
21	8,00	53	8,00	85	9,50	117	10,00
22	9,25	54	8,50	86	5,50	118	8,00
23	8,25	55	9,00	87	10,00	119	8,25
24	8,75	56	8,00	88	5,00	120	7,75
25	5,75	57	7,00	89	5,75	121	8,50
26	8,00	58	8,50	90	6,50	122	8,75
27	7,75	59	7,75	91	9,25	123	9,00
28	8,50	60	8,25	92	6,00	124	7,75
29	5,50	61	8,00	93	9,50	125	8,00
30	8,25	62	8,75	94	5,50		
31	7,50	63	7,50	95	10,00		
32	7,00	64	9,00	96	9,00		

T3. Corrélation

3.1 Les accidents de la circulation au Québec

3.2 Les échelles salariales du personnel enseignant

3.1 Les accidents de la circulation au Québec

Le tableau 3.1 donne pour chaque région administrative du Québec, le total des accidents de la circulation et le nombre d'accidents mortels sur les routes en 1978. Ces chiffres ont été recueillis par les services de police municipaux et la Sûreté du Québec.

Tableau 3.1 Les accidents de la circulation par région administrative - 1978.

Région administrative	Total des accidents	Accidents mortels
Bas St-Laurent	7 636	99
Saguenay	12 957	77
Québec	41 377	238
Mauricie	15 309	142
Cantons de l'Est	8 670	64
Montréal	137 027	515
Outaouais	10 890	67
Nord-Ouest	6 285	54
Côte-Nord	5 140	39
Nouveau-Québec	757	9

Source: Rapport de la Commission de police du Québec, 1978

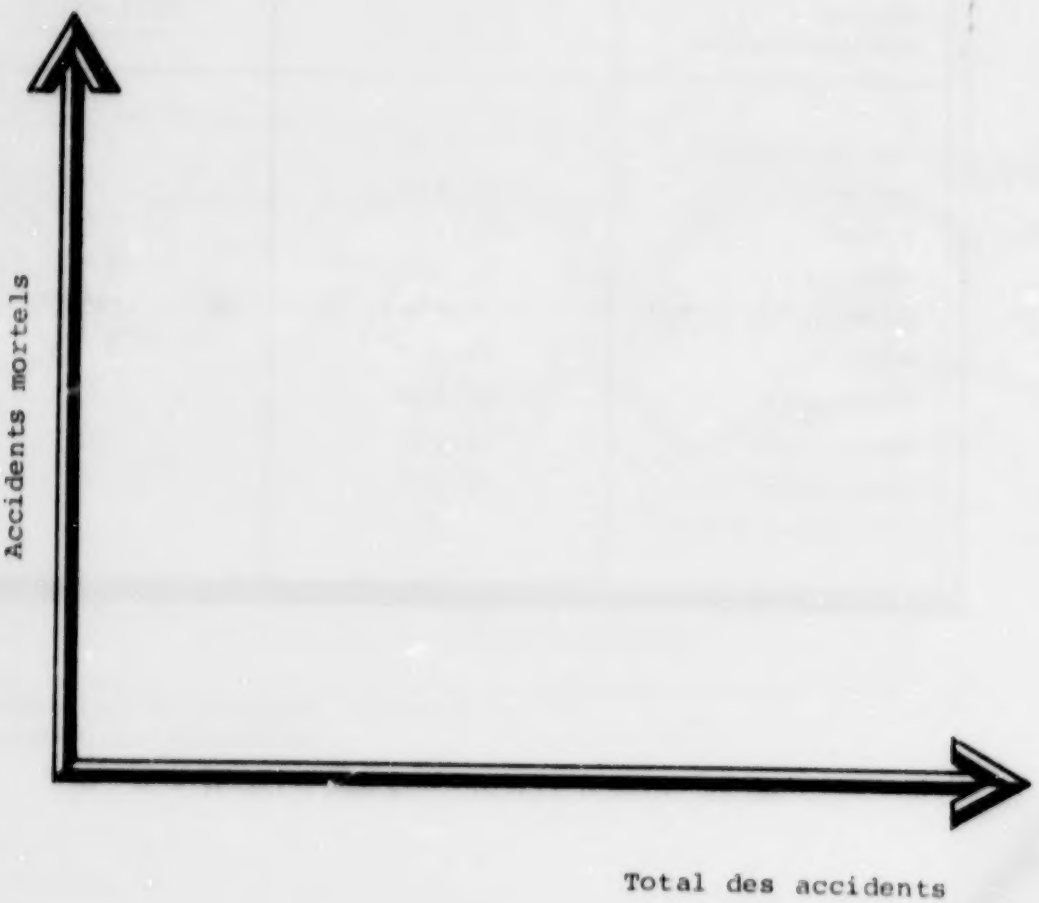
1) L'existence d'une relation entre le nombre total d'accidents et le nombre d'accidents mortels.

Pour vérifier l'existence d'une relation, il y a différents moyens.

a) Moyen visuel

Faire une représentation graphique entre le nombre total d'accidents et le nombre d'accidents mortels.

Figure 3.1 Le nombre d'accidents mortels en fonction du nombre total d'accidents



b) Moyen analytique

Calculer le coefficient de corrélation linéaire entre les deux caractéristiques

r = ...

Voir le manuel d'instructions au sujet de cet instrument.

2) Le genre de la relation

Déterminer le genre de relation entre les deux caractéristiques à partir du graphique déjà fait.

3) L'expression mathématique de la relation

Nombre d'accidents mortels = m (nombre total d'accidents) + b

Trouver m et b

m = ...

b = ...

Voir le manuel d'instructions.

Tracer la droite à la figure 3.1.

Commenter la situation du Bas St-Laurent par rapport à l'ensemble du Québec.

3.2 Les échelles salariales du personnel enseignant

3.2.1 Le tableau 3.2 donne le nombre d'année(s) d'expérience et le salaire hebdomadaire correspondant pour les professeurs de Cégep qui ont seize ans de scolarité.

Tableau 3.2 Le nombre d'année(s) d'expérience et le salaire hebdomadaire des professeurs de Cégep ayant 16 ans de scolarité (1981-82)

Nombre d'année(s) d'expérience	Salaire hebdomadaire (arrondi au $ près)
0	
1	375
2	387
3	399
4	412
5	425
6	439
7	454
8	469
9	484
10	500
11	515
12	533
13	550
14	568
	586

1) L'existence d'une dépendance linéaire entre le nombre d'année(s) d'expérience et le salaire hebdomadaire.

a) Moyen visuel

Faire le graphique du salaire hebdomadaire en fonction du nombre d'année(s) d'expérience.

Figure 3.2 Le salaire hebdomadaire en fonction du nombre d'année(s) d'expérience

D'après le graphique, déterminer le genre et le degré de la dépendance linéaire entre les deux caractéristiques.

b) Moyen analytique

Calculer le coefficient de corrélation linéaire (r) entre les deux caractéristiques

r = ...

D'après la valeur de r, déterminer le genre et le degré de la dépendance linéaire entre les deux caractéristiques. Comparer cette réponse avec celle trouvée par le moyen visuel.

2) La recherche d'un modèle pour exprimer la dépendance linéaire entre les deux caractéristiques

Trouver l'équation de la droite des moindres carrés.

Salaire hebdomadaire =

m(nombre d'année(s) d'expérience) + b

Tracer la droite à la figure 3.2.

D'après ce modèle, trouver le salaire hebdomadaire d'un professeur ayant 8 années d'expérience.

3.2.2 Le tableau 3.3 donne le nombre d'années de scolarité et le salaire hebdomadaire correspondant pour les professeurs de Cégep qui ont cinq ans d'expérience.

Tableau 3.3 Le nombre d'années de scolarité et le salaire hebdomadaire des professeurs de Cégep ayant 5 ans d'expérience (1981-82)

Nombre d'année(s) de scolarité	Salaire hebdomadaire (arrondi au $ près)
16	439
17	476
18	515
19	589

1) L'existence d'une dépendance linéaire entre le nombre d'années de scolarité et le salaire hebdomadaire.

a) Moyen visuel

Faire le graphique du salaire hebdomadaire en fonction du nombre d'années de scolarité.

Déterminer le genre et le degré de la dépendance linéaire entre les deux caractéristiques.

b) Moyen analytique

Calculer le coefficient de corrélation linéaire entre les deux caractéristiques.

r = ...

2) La recherche d'un modèle pour exprimer la dépendance linéaire entre les deux caractéristiques.

Trouver l'équation de la droite des moindres carrés

Salaire hebdomadaire =

m(nombre d'années de scolarité) + b

D'après ce modèle, trouver le salaire hebdomadaire d'un professeur ayant 16 années de scolarité.

Q 1. *Est-ce que l'équation trouvée en 3.2.1 permet de calculer le salaire hebdomadaire d'un professeur de façon plus précise que celle trouvée en 3.2.2?*

Justifier.

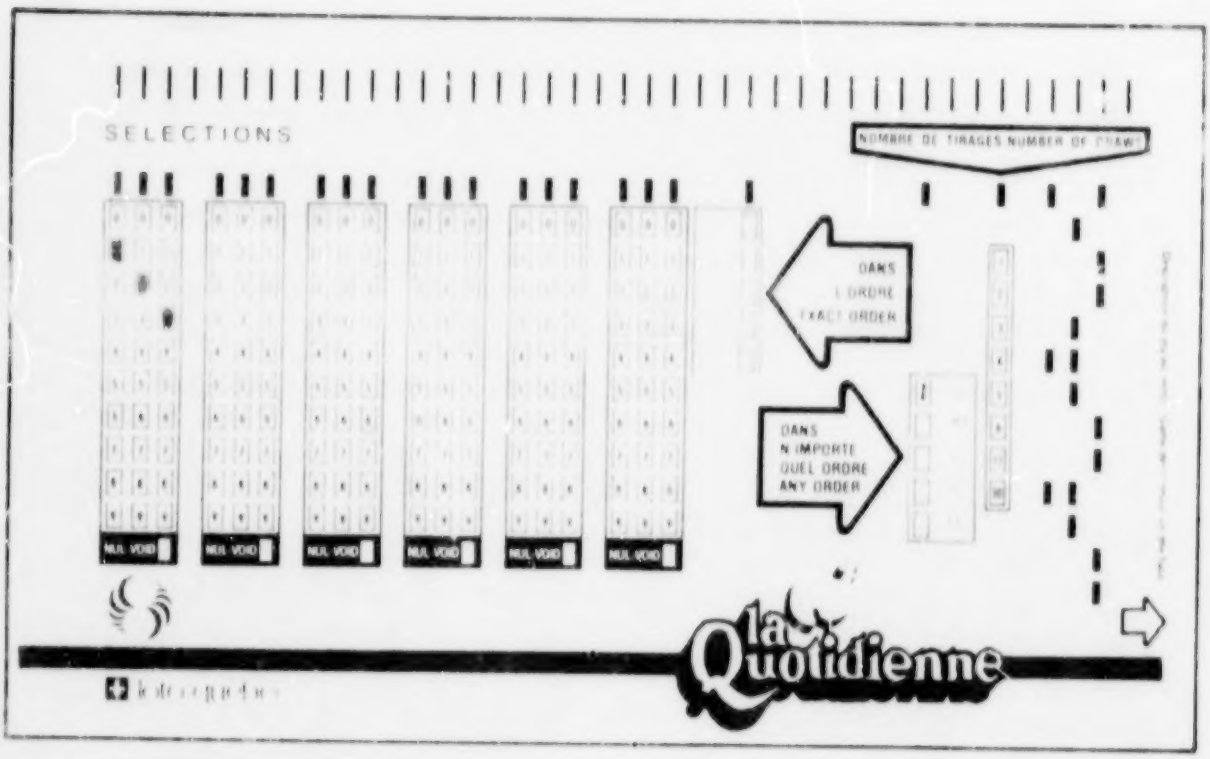
T4. Probabilité

4.1 Le jeu de La Quotidienne

4.2 Le jeu de la mini-loto

4.3 La règle des parties

4.1 Le jeu La Quotidienne



À La Quotidienne, pour une mise de 0,50 $, tu choisis un numéro entre 000 et 999 inclusivement. Tu peux jouer ce numéro soit dans l'ordre soit dans n'importe quel ordre. Loto-Québec tire au hasard un numéro gagnant de trois chiffres.

Tu peux gagner de deux façons, 450 fois ta mise, soit 225 $: 1) si tu joues dans l'ordre et si le numéro de ton billet est dans le même ordre que le numéro gagnant, 2) si tu joues dans n'importe quel ordre et si ton numéro tout comme le numéro gagnant comprend trois chiffres identiques.

Tu gagnes 150 fois ta mise, soit 75 $, si tu joues dans n'importe quel ordre, si tu choisis un numéro de deux chif-

fres identiques et si ton numéro contient les mêmes chiffres que ceux du numéro gagnant.

Tu gagnes 75 fois ta mise, soit 37,50 $, si tu joues dans n'importe quel ordre, si tu choisis un numéro de trois chiffres différents et si ton numéro possède les mêmes chiffres que le numéro gagnant.

Si tu achètes un billet,

Q 1. quelle est la probabilité de gagner 225 $, 75 $, 37,50 $ et 0 $ si

i) tu choisis de jouer dans l'ordre?

ii) tu choisis de jouer dans n'importe quel ordre avec un numéro de trois chiffres identiques?

iii) tu choisis de jouer dans n'importe quel ordre avec un numéro de deux chiffres identiques et l'autre différent?

iv) tu choisis de jouer dans n'importe quel ordre avec un numéro de trois chiffres différents?

Q 2. quel est ton gain net à chaque fois?

Un jeu de hasard est équitable lorsque l'espérance mathématique de gain de chaque joueur est nulle. L'espérance mathématique est définie de la sorte: supposons que $x_1, x_2, \ldots, x_n$ sont les gains possibles d'un joueur et que $Pr(x_1), Pr(x_2), \ldots, Pr(x_n)$ sont les probabilités respectives de gagner ces montants, alors l'espérance mathématique de ce joueur, notée E(X), est donnée par:

$$E(X) = x_1Pr(x_1) + x_2Pr(x_2) + \ldots + x_nPr(x_n).$$

L'espérance de gain représente le gain moyen de quelqu'un qui joue un grand nombre de fois.

Q 3. Quelle est l'espérance mathématique du gain net pour chacune des quatre situations décrites à Q 1.?

Q 4. Parmi les quatre situations décrites à Q 1., y en a-t-il une plus avantageuse que les autres?

Q 5. As-tu encore espoir de faire fortune en prenant régulièrement des billets de La Quotidienne?

4.2 Le jeu de la Mini-Loto

À la Mini-loto, les billets son numérotés de 100 000 à 999 999 inclusivement. Afin de déterminer le numéro gagnant, un nombre est choisi au hasard à chaque semaine. Le détenteur d'un billet gagne 50 000 $ si le numéro de son billet est le même que le numéro gagnant, 5 000 $ si les 5 derniers chiffres de son billet correspondent dans l'ordre aux 5 derniers chiffres du numéro gagnant, 250 $ si les 4 derniers chiffres de son billet correspondent dans l'ordre aux 4 derniers chiffres du numéro gagnant, 50 $ si les 3 derniers chiffres de son billet correspondent dans l'ordre aux 3 derniers chiffres du numéro gagnant et de 5 $ si les 2 derniers chiffres de son billet correspondent aux 2 derniers chiffres du numéro gagnant.
Si tu achètes un billet la semaine prochaine (0,50 $),

Q 1. quelle est la probabilité de gagner 50 000 $, 5 000 $, 250 $, 50 $, 5 $ et 0 $ respectivement?

Q 2. *quel est ton gain net pour chaque cas énuméré à Q 1.?*

Q 3. *quelle est l'espérance mathématique de ton gain net avec un billet de Mini-loto?*

Q 4. *penses-tu qu'il est préférable de jouer régulièrement à la Mini-loto plutôt qu'à la Quotidienne?*

4.3 La règle des parties

Le début des probabilités remonte au 17e siècle lorsque le chevalier de Merée posa à son ami Pascal, le problème suivant: si deux joueurs jouent en 3 parties (c'est-à-dire que le gagnant est celui qui, le premier totalise 3 parties gagnées) et s'ils désirent interrompre prématurément le jeu, comment les mises doivent-elles être réparties entre eux, compte tenu des parties déjà gagnées par l'un et l'autre? (On suppose que les joueurs jouent à pile ou face.)

Déterminer le partage à faire entre deux joueurs ayant misé chacun 32 $ et qui s'arrêtent lorsque

1) les deux ont gagné chacun 2 parties;

2) l'un a gagné 2 parties et l'autre 1;

3) l'un a gagné 2 parties et l'autre 0;

4) l'un a gagné 1 partie et l'autre 0.

Formalisation

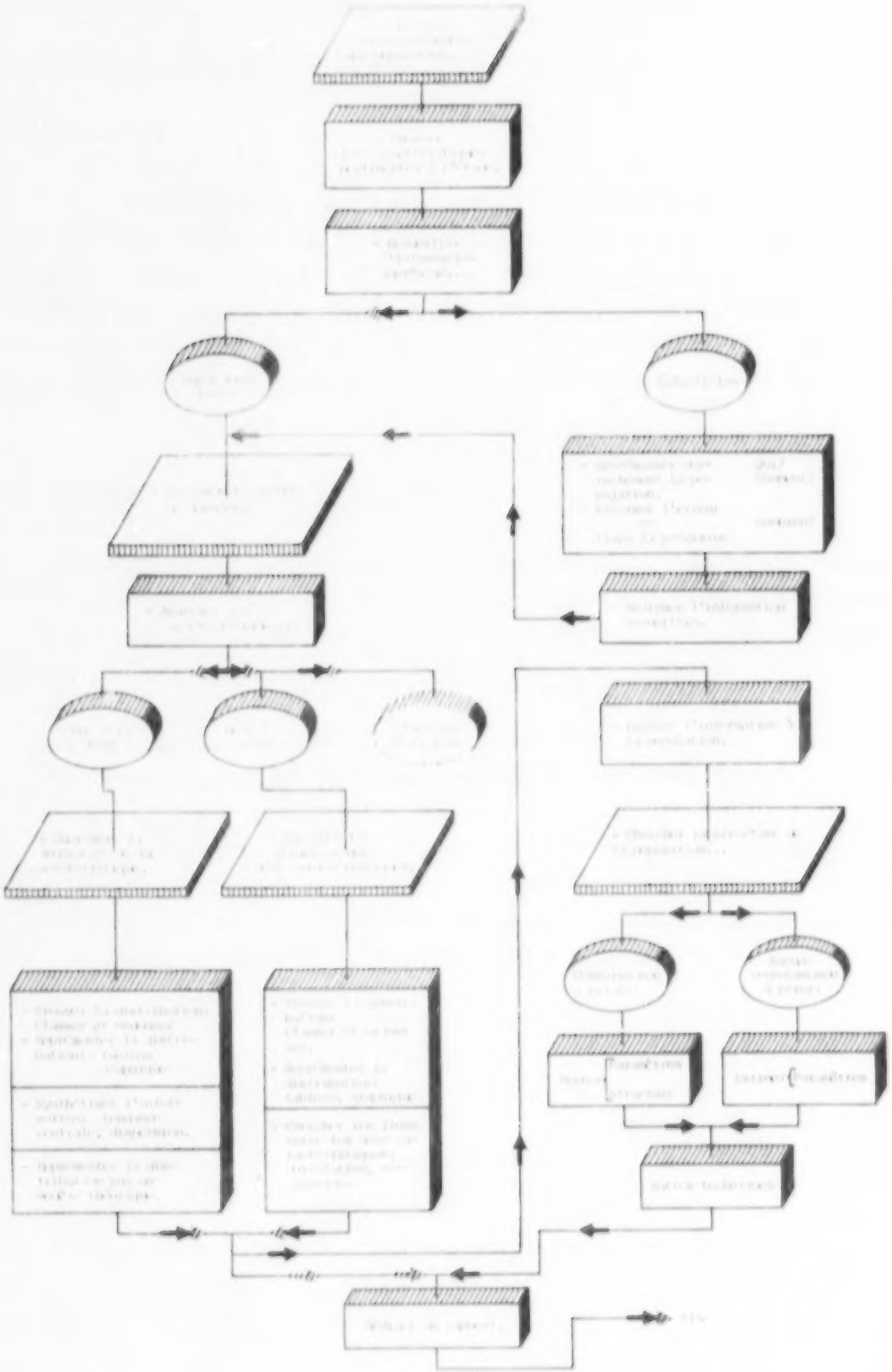
OBJECTIF GÉNÉRAL

Prendre conscience de la valeur de la formalisation, qui a pour but de synthétiser, de résumer, d'exprimer rigoureusement les différents concepts des probabilités et de la statistique et d'établir des liens entre eux.

CONTEXTE DE LA FORMALISATION

Au laboratoire expérimental, tu as assimilé de façon intuitive un certain nombre de concepts des probabilités et de la statistique. Ces concepts ont été définis dans les textes de symbolisation et tu les as utilisés au laboratoire d'applications. Dans la formalisation, nous voulons résumer tous les concepts vus jusqu'à maintenant, les synthétiser, les exprimer rigoureusement et établir des liens entre eux.

Description statistique d'une population



Description statistique d'une population

Introduction

Pour décrire statistiquement une population, il faut bien définir l'objet de l'étude, puis choisir les caractéristiques pertinentes à cette étude. Ensuite, on doit recueillir l'information auprès de la population entière ou seulement une partie de la population, c'est-à-dire un échantillon. Dans les deux cas, on devra analyser le groupe de données (voir le schéma).

1. Analyse d'un groupe de données

Pour analyser un groupe de données, on doit étudier les caractéristiques retenues au début de l'étude. Les caractéristiques peuvent être analysées une à la fois, deux à la fois ou plusieurs à la fois.

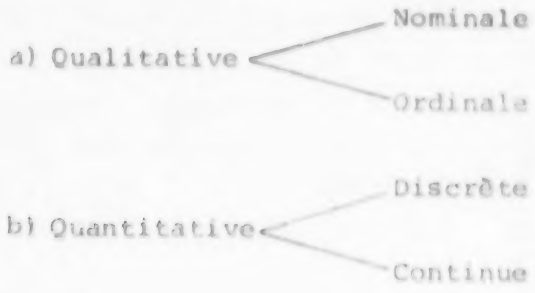
1.1 Analyse d'une caractéristique à la fois (recherche de la structure de la caractéristique)

Rechercher la structure d'une caractéristique signifie qu'il faut trouver sa distribution, en faire une représentation, la synthétiser et l'exprimer, si possible, par un modèle théorique ou une loi mathématique.

1.1.1 Nature d'une caractéristique

Il faut trouver les catégories de chaque caractéristique, c'est-à-dire une liste complète des subdivisions mutuellement exclusives.

La nature d'une caractéristique peut être:



Une caractéristique est:

a) Qualitative
si ses catégories ne proviennent ni d'une mesure ni d'un dénombrement,

1) Nominale
si ses catégories ne peuvent être ordonnées; par exemple, masculin et féminin pour le sexe,

2) Ordinale
si ses catégories peuvent être ordonnées; par exemple, collégial I, collégial II, collégial III pour le niveau scolaire;

b) Quantitative
si ses catégories proviennent d'une mesure ou d'un dénombrement; on l'appelle aussi variable statistique,

1) Discrète
si ses catégories ne sont que des valeurs isolées; par exemple, le nombre d'enfants par famille,

2) Continue
si, à priori, ses catégories peuvent être des valeurs quelconques dans un intervalle donné; par exemple, la taille des personnes.

Il est important de connaître la nature des caractéristiques car les outils pour les analyser peuvent être différents.

1.1.2 Distribution d'une caractéristique

Après avoir déterminé la nature de la caractéristique étudiée, il faut classifier et ordonner, si possible, les résultats obtenus. Il faut trouver l'effectif ou la fréquence de chaque catégorie de la caractéristique. La distribution qui en résulte peut être représentée dans un tableau à simple entrée et/ou un graphique approprié.

1.1.3 Représentation de la distribution d'une caractéristique

1.1.3.1 Tableau à simple entrée

a) Caractéristique nominale

Il s'agit d'énumérer les différentes catégories dans une colonne puis les effectifs ou les fréquences dans une autre colonne. On donne l'effectif (ou la fréquence) total, et on indique, s'il y a lieu, un titre, une date, et la provenance des données.

Tableau 1 ----------

Caractéristique	Effectif
Catégorie 1	----------
Catégorie 2	----------
----------	----------
----------	----------
Total	----------

Source: ----------

Les fréquences peuvent être exprimées sous forme de rapports, de fractions décimales ou de pourcentages.

b) Caratéristique ordinale

On ordonne les différentes catégories puis on procède de la même façon qu'en a).

c) Caractéristique continue

On peut procéder comme dans le cas d'une caractéristique ordinale. Toutefois, il arrive souvent que le nombre de catégories différentes est très élevé. Il convient alors de regrouper un certain nombre de catégories différentes dans une même classe (voir le manuel d'instructions à ce sujet). Le tableau suivant en est un exemple:

Tableau 2 La répartition des poids d'un groupe de 100 personnes

Poids (kg)	Nombre de personnes
60 - 62	8
63 - 65	20
66 - 68	39
69 - 71	27
72 - 74	6
Total	100

Si les poids sont mesurés à un kilogramme près, la classe 60-62 comprend théoriquement tous les poids variant de 59,5 kg à 62,5 kg (exclusivement). Les valeurs 59,5 et 62,5 sont appelées les *bornes de la classe*. Les valeurs 60 et 62 sont appelées les *limites de la classe*. La *largeur d'une classe* est la différence entre les deux bornes de la classe. Le *centre d'une classe* est le point milieu de cette classe et il s'obtient en divisant par deux la somme des limites de la classe.

Une caractéristique continue peut aussi être représentée dans un tableau des effectifs (ou des fréquences) cumulés. L'effectif cumulé (ou la fréquence cumulée) d'une classe est l'effectif total (ou la fréquence totale) correspondant à toutes les valeurs plus petites que la borne supérieure de la classe.

Tableau 3 La répartition des poids d'un groupe de 100 personnes

Poids (kg)	Nombre de personnes
moins de 59,5	0
moins de 62,5	8
moins de 65,5	28
moins de 68,5	67
moins de 71,5	94
moins de 74,5	100

d) Caractéristique discrète

Si les catégories différentes ne sont pas très nombreuses, on les ordonne puis on construit un tableau à simple entrée. Sinon, on regroupe les données en classes, comme c'est le cas pour une caractéristique continue.

1.1.3.2 Graphiques

Voici un résumé des conventions à observer dans la construction des différents graphiques.

Résumé

CONSTRUCTION D'UN GRAPHIQUE

Sorte de graphique	Nature de la caractéristique	Abcisse	Ordonnée	Description
Diagramme en bâtons	Qualitative ou discrète	Différentes catégories	Effectifs ou fréquences	Un ensemble de bâtons
Histogramme	Continue	Bornes ou centres de classe	Effectifs ou fréquences	Ensemble de rectangles contigus
Polygone des effectifs (fréquences)	Continue	Centres de classe	Effectifs ou fréquences	Polygone
Polygone des effectifs (ou fréquences) cumulés	Continue	Bornes de classe	Effectifs ou fréquences cumulés	Polygone

Résumé

Distribution d'une caractéristique et représentation

1. Nature d'une caractéristique

   Qualitative - Nominale
   - Ordinale

   Quantitative - Discrète
   - Continue

2. Représentation de la distribution d'une caractéristique

   a) Tableau à simple entrée

      Énumérer les différentes catégories dans une colonne puis les effectifs dans une autre colonne.

   b) Graphiques

      1. Diagramme à bâtons

      2. Histogramme

      3. Polygone des effectifs (fréquences)

      4. Polygone des effectifs (ou fréquences) cumulés

1.1.4 Synthèse de l'information

Après avoir trouvé la distribution d'une caractéristique et l'avoir représentée, il peut être intéressant de condenser davantage l'information recueillie auprès du groupe dans le but de la comparer à celle provenant d'un autre groupe. Il convient alors d'évaluer la tendance centrale et la dispersion du groupe de données.

1.1.4.1 Tendance centrale

a) Définition

C'est une valeur représentative de l'ensemble des données c'est-à-dire une valeur qui donne une idée sommaire de l'ordre de grandeur des données.

b) Outils pour mesurer la tendance centrale

1) Moyenne arithmétique ($\bar{x}$ ou μ)

i) Définition

La valeur obtenue suite à une redistribution uniforme de la somme des données.

$$\text{Moyenne arithmétique} = \frac{\sum_{i=1}^{p} x_i}{p} \qquad (1)$$

où x_i : la $i^{ième}$ donnée individuelle

$\sum$: indique la somme

p : le nombre de données

ii) Procédés de calcul

1. Lorsqu'il y a plusieurs données identiques, les calculs engendrés par la formule (1) sont simplifiés si on effectue:

$$\frac{\sum_{i=1}^{nc} e_i \, c_i}{\sum_{i=1}^{nc} e_i} \qquad (2)$$

où c_i: valeur de la $i^{ième}$ catégorie

e_i: effectif de la $i^{ième}$ catégorie

nc: nombre de catégories.

2. La formule (2) peut aussi s'exprimer en fonction de la fréquence de chaque catégorie.

$$= \sum_{i=1}^{nc} f_i \, c_i \qquad (3)$$

3. Si les données ont été regroupées en classes, on ne peut trouver qu'une approximation de la moyenne arithmétique. On calcule alors:

$$\frac{\sum_{i=1}^{ncl} e_i \, m_i}{\sum_{i=1}^{ncl} e_i} = \sum_{i=1}^{ncl} f_i \, m_i$$

où m_i: centre de la $i^{ième}$ classe

n:l: nombre de classes.

iii) Propriétés

Théorème 1: *Soient une variable X et une variable Y, si Y = aX + b alors m.a.de Y = a(m.a.de X) + b où a et b sont des constantes.*

Preuve:

$$\text{m.a. de } Y = \frac{Y_1 + Y_2 + \text{---} + Y_p}{p}$$

$$= \frac{(ax_1+b)+(ax_2+b)+ \text{---} + (ax_p+b)}{p}$$

$$= \frac{a(x_1+x_2+ \text{---} + x_p) + pb}{p}$$

$$= a\left(\frac{x_1+x_2+ \text{---} + x_p}{p}\right) + b$$

$$= a(\text{m.a. de } X) + b$$

Théorème 2: *Si une variable X prend les valeurs $x_1, x_2, \ldots, x_p$ alors la somme des déviations de chaque valeur par rapport à la moyenne arithmétique de X est nulle i.e.*

$$\sum_{i=1}^{p} (x_i - \text{m.a. de } X) = 0$$

Preuve:

$$\sum_{i=1}^{p} (x_i - \text{m.a. de X}) = \sum_{i=1}^{p} x_i - \sum_{i=1}^{p} \text{m.a.de X}$$

$$= \sum_{i=1}^{p} x_i - p(\text{m.a. de X})$$

$$= \sum_{i-1}^{p} x_i - p\left(\frac{\sum_{i=1}^{p} x_i}{p}\right)$$

$$= \sum_{i=1}^{p} x_i - \sum_{i=1}^{p} x_i$$

$$= 0$$

2) Mode (mo)

La(les) valeur(s) la(les) plus fréquente(s).

Si les données ont été regroupées en classe, on peut seulement trouver la classe modale c'est-à-dire la classe dont l'effectif ou la fréquence est relativement le(la) plus élevé(e).

3) Médiane (md)

C'est une valeur telle qu'au moins la moitié des données ordonnées sont plus petites ou égales et au moins la moitié des données sont plus grandes ou égales à cette valeur.

c) Outils pour mesurer la position

1) Quartiles (Q_1, Q_2, Q_3)

Ce sont trois valeurs qui, dans une suite de nombres ordonnés, partagent l'effectif total en quatre parties de même effectif.

On remarque que Q_2 = md

Le graphique suivant illustre une façon de trouver approximativement la valeur de chaque quartile.

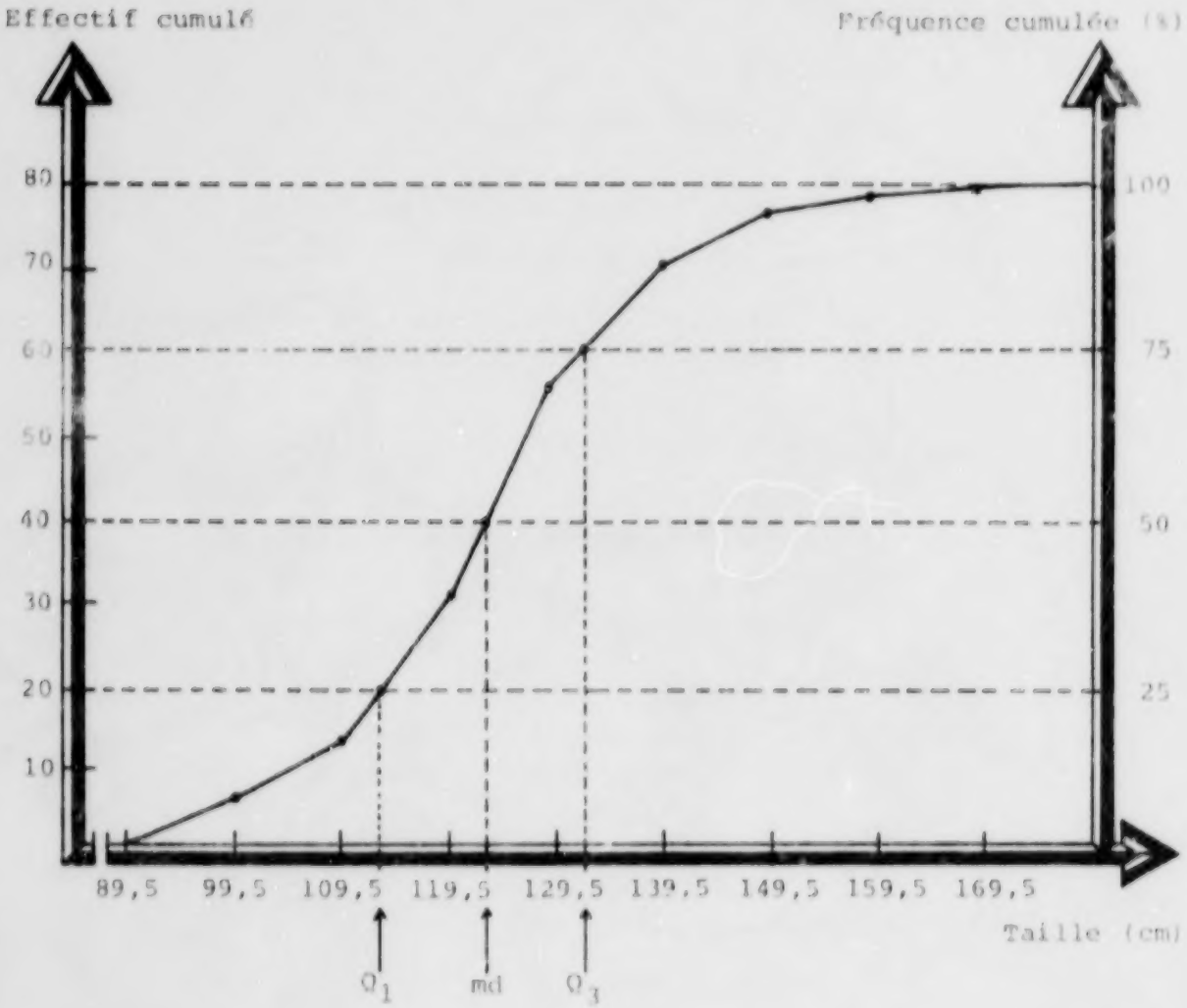


Figure 2 Illustration des quartiles

2) Centiles

Ce sont quatre-vingt-dix-neuf (99) valeurs qui, dans une suite de nombres ordonnés, partagent l'effectif total en cent (100) parties de même effectif.

Le rang centile d'une donnée est le pourcentage de données qui lui sont inférieures.

1.1.4.2 Dispersion

a) Définition

C'est la variation des données.

b) Outils pour mesurer la dispersion.

1) Étendue simple (R)

C'est la différence entre la plus grande et la plus petite donnée d'un groupe.

2) Écart moyen (E.M.)

L'écart moyen est la moyenne des écarts entre chaque donnée et la moyenne arithmétique des données. (L'écart entre deux valeurs est la valeur absolue de leur différence).

$$\text{E.M.} = \frac{\sum_{i=1}^{p} | x_i - \text{m.a. de } X |}{p}$$

3) Variance (s^2 ou σ^2)

i) Définition

C'est la moyenne des carrés des écarts entre chaque donnée et la moyenne arithmétique des données.

$$\text{Var.} = \frac{\sum_{i=1}^{p} (x_i - \text{m.a. de } X)^2}{p} \qquad (1)$$

ii) Procédés de calcul

1. Comme nous l'avons vu pour la moyenne, s'il y a plusieurs données identiques, il est plus rapide d'effectuer les calculs suivant:

$$\frac{\sum_{i=1}^{nc} (e_i)(c_i - \text{m.a. de } X)^2}{\sum_{i=1}^{nc} e_i} \qquad (2)$$

ou

$$\sum_{i=1}^{nc} (f_i)(c_i - \text{m.a. de } X)^2 \qquad (3)$$

2. Si les données sont regroupées en classes de centre m_i, on trouve une approximation de la variance en remplaçant dans les formules (2) et (3) c_i par m_i.

3. Lorsque la moyenne arithmétique est fractionnaire, il est plus rapide de calculer la variance en modifiant la formule (1) pour obtenir:

$$\frac{\sum_{i=1}^{p} x_i^2}{p} - \left(\frac{\sum_{i=1}^{p} x_i}{p}\right)^2 \qquad (4)$$

Les formules (1) et (4) sont équivalentes.

Preuve:

Soient x_1, x_2, ---, x_p les p valeurs de la variable X alors:

var. de X

$$= \frac{\sum_{i=1}^{p} (x_i - \text{m.a. de } X)^2}{p}$$

$$= \frac{\sum_{i=1}^{p} \left[x_i^2 - 2x_i (\text{m.a. de } X) + (\text{m.a. de } X)^2 \right]}{p}$$

$$= \frac{\sum_{i=1}^{p} x_i^2 - \sum_{i=1}^{p} 2x_i (\text{m.a. de } X) + \sum_{i=1}^{p} (\text{m.a. de } X)^2}{p}$$

$$= \frac{\sum_{i=1}^{p} x_i^2 - 2(\text{m.a. de } X) \sum_{i=1}^{p} x_i + p(\text{m.a. de } X)^2}{p}$$

$$= \frac{\sum_{i=1}^{p} x_i^2 - \frac{\sum_{i=1}^{p} x_i \sum_{i=1}^{p} x_i}{p} + p\left(\frac{\sum_{i=1}^{p} x_i}{p}\right)^2}{p}$$

$$= \frac{\sum_{i=1}^{p} x_i^2}{p} - 2\left(\frac{\sum_{i=1}^{p} x_i}{p}\right) + \left(\frac{\sum_{i=1}^{p} x_i}{p}\right)^2$$

$$= \frac{\sum_{i=1}^{p} x_i^2}{p} - \left(\frac{\sum_{i=1}^{p} x_i}{p}\right)^2$$

Cette dernière formule est utilisée dans la programmation des calculatrices ou des ordinateurs.

4. La variance peut être calculée de la façon suivante:

$$\frac{\sum_{i=1}^{nc} (e_i)(c_i^2)}{\sum_{i=1}^{nc} e_i} - \left(\frac{\sum_{i=1}^{nc} (e_i)(c_i)}{\sum_{i=1}^{nc} e_i}\right)^2$$

ou

$$\sum_{i=1}^{nc} (f_i)(c_i^2) - \left(\sum_{i=1}^{nc} (f_i)(c_i)\right)^2$$

5. Lorsque les données sont regroupées en classes, il s'agit de remplacer c_i par m_i dans les deux dernières formules.

iii) Propriétés

Théorème 3: *Soient une variable X et une variable Y, si* $Y = aX + b$ *alors* $\text{var. de } Y = a^2 \text{ var. de } X$ *où a et b sont des constantes*

Preuve:

$$\text{var. de } Y = \frac{\sum_{i=1}^{p} (y_i - \text{m.a. de } Y)^2}{p}$$

$$= \frac{\sum_{i=1}^{p} \left[ax_i + b - (a(\text{m.a. de } X) + b)\right]^2}{p}$$

(d'après le théorème 1)

$$= \frac{\sum_{i=1}^{p} \left[a(x_i - \text{m. de } X)\right]^2}{p}$$

$$= \frac{\sum_{i=1}^{p} a^2 (x_i - \text{m.a. de } X)^2}{p}$$

$$= \frac{a^2 \sum_{i=1}^{p} (x_i - \text{m.a. de } X)^2}{p}$$

$$= a^2 \text{ var. de } X$$

Corollaire 1: Si à chaque valeur d'une variable on additionne une constante, la variance ne change pas.

Preuve: On pose a = 1 dans le théorème précédent. On obtient alors

$$Y = X + b$$

et

$$\text{var. de } Y = \text{var. de } X.$$

4) Écart-type (s ou σ)

i) Définition

C'est la racine carrée de la variance.

ii) Propriétés

Théorème 4: *Soient deux variables X et Y liées par l'équation*

$$Y = aX + b$$

alors

$$\text{é.-t. de } Y = |a| \text{ é.-t. de } X.$$

Preuve: La conclusion de ce théorème découle du théorème 3 et de la définition de l'écart-type.

Théorème 5. *Soit la variable Z, appelée aussi cote Z définie par rapport à une variable X de la façon suivante:*

$$Z = \frac{X - \text{m.a. de } X}{\text{é.-t. de } X}$$

alors

1) m.a. de Z = 0

2) é.-t. de Z = 1

Preuve:

Montrons d'abord que Z est une transformation de X.

$$Z = \frac{X - \text{m.a. de } X}{\text{é.-t. de } X}$$

$$Z = \left(\frac{1}{\text{é.-t. de } X}\right) X + \left(-\frac{\text{m.a. de } X}{\text{é.-t. de } X}\right)$$

où $\frac{1}{\text{é.-t. de } X}$ et $\left(-\frac{\text{m.a. de } X}{\text{é.-t. de } X}\right)$

sont des constantes.

1) $$\text{m.a. de } Z = \left(\frac{1}{\text{é.-t. de } X}\right)(\text{m.a. de } X) + \left(-\frac{\text{m.a. de } X}{\text{é.-t. de } X}\right)$$

(Théorème 1)

m.a. de Z = 0

2) $$\text{é.-t. de } Z = \left|\frac{1}{\text{é.-t. de } X}\right| (\text{é.-t. de } X)$$

(Théorème 5)

é.-t. de Z = 1.

Résumé

Synthèse de l'information d'une caractéristique

1. Tendance centrale

Outils pour mesurer la tendance centrale.

a) Moyenne arithmétique (μ ou $\bar{x}$) = $\dfrac{\sum_{i=1}^{p} x_i}{p}$

b) Mode (mo)

La(les) valeur(s) la(les) plus fréquente(s)

c) Médiane (md)

Valeur telle qu'au moins la moitié des données ordonnées sont plus petites ou égales et au moins la moitié des données sont plus grandes ou égales à cette valeur.

2. Dispersion

Outils pour mesurer la dispersion.

a) Etendue simple (R)

La différence entre la plus grande et la plus petite donnée d'un groupe.

b) Écart moyen (E.M.)

$$= \frac{\sum_{i=1}^{p} \left| x_i - \text{m.a. de } X \right|}{p}$$

c) Variance (s^2 ou σ^2)

$$= \frac{\sum_{i=1}^{p} (x_i - \text{m.a. de } X)^2}{p}$$

d) Écart-type (s ou σ)

C'est la racine carrée de la variance.

3. Propriétés

Soient une variable X et une variable Y.

Si $Y = aX + b$,

a) alors m.a. de $Y = a$(m.a. de X) + b

b) alors var. de $Y = a^2$ var. de X

c) alors é.-t. de $Y = |a|$ é.-t. de X

où a et b sont des constantes.

Par exemple:

$$Z = \frac{X - \text{m.a. de } X}{\text{é.-t. de } X}$$

alors

1) m.a. de Z = 0

2) é.-t. de Z = 1.

1.1.5 Modèles théoriques de l'information

1.1.1.5 Modèle pour les variables discrètes

Loi binômiale B(n;p)

1) Description de la structure

i) Expression analytique

$$f(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$0 \leq k \leq n$

n et k des entiers non négatifs

$$\binom{n}{k} = \frac{n!}{k!\ (n-k)!}$$

$n! = n(n-1)(n-2)\ \cdots\ (2)(1)$

$0! = 1$

ii) Caractéristiques

Les caractéristiques varient en fonction de p.

	$p < \frac{1}{2}$	$p = \frac{1}{2}$	$p > \frac{1}{2}$
Pic - nombre - localisation Symétrie par rapport au pic	1 à gauche non	1 au centre oui	1 à droite non

2) Synthèse de la structure

Moyenne de X = np

Variance de X = np(1-p)

1.1.5.2 Modèles pour les variables continues

a) Loi uniforme

1) Description de la structure

i) Expression analytique

$$f(x) = \frac{1}{b - a} \qquad a < x < b$$

ii) Caractéristiques

- aucun pic
- symétrie par rapport à la tendance centrale

2) Synthèse de la structure

$$\text{Moyenne de } X = \frac{a + b}{2}$$

$$\text{Variance de } X = \frac{(b - a)^2}{12}$$

b) Loi normale $\mathcal{N}$ (moyenne; écart-type)

1) Description de la structure

i) Expression analytique

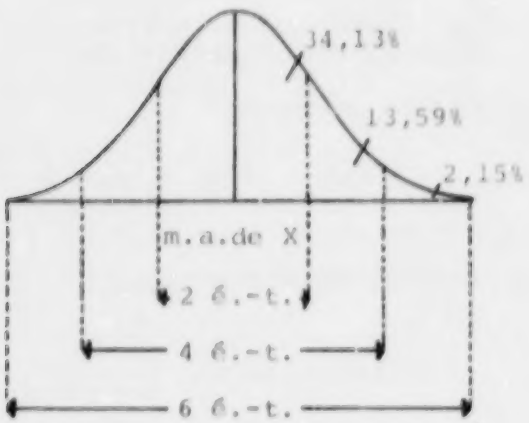
$$f(x) = \frac{1}{\sqrt{2\pi}\ (\text{é.-t. de X})}\ e^{-\frac{1}{2}\left(\frac{x - \text{m.a. de X}}{\text{é.-t. de X}}\right)^2}$$

où $\pi = 3,14159...$

$e = 2,718...$

ii) Caractéristiques

- un pic central
- symétrie par rapport à ce pic central
- courbe en forme de cloche
- pourcentage



2) Synthèse de la structure

Moyenne de X = μ ou $\bar{x}$

Variance de X = σ^2 ou s^2

Résumé

Modèles théoriques de l'information

1. Modèle pour les variables discrètes

   Loi binômiale $B(n;p)$

   $$f(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \leq k \leq n$$

   n et k des entiers non négatifs

   $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

   Moyenne de $X = np$

   Variance de $X = np(1 - p)$

2. Modèles pour les variables continues

   a) Loi uniforme

   $$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

   Moyenne de $X = \frac{a + b}{2}$

   Variance de $X = \frac{(b - a)^2}{12}$

   b) Loi normale $\mathcal{N}$(moyenne; écart-type)

   $$f(x) = \frac{1}{\sqrt{2\pi}\ (\text{é.-t. de } X)} e^{-\frac{1}{2}\left(\frac{x - \text{m.a. de } X}{\text{é.-t. de } X}\right)^2}$$

   $-\infty < x < +\infty$

   Moyenne de $X = \mu$ ou $\bar{x}$

   Variance de $X = \sigma^2$ ou s^2

1.2 Analyse de deux caractéristiques à la fois (recherche de la structure conjointe des deux caractéristiques)

Sans oublier ce qui vient d'être vu au sujet d'une caractéristique, nous allons étudier le fait de prendre deux caractéristiques à la fois et cela, autant pour les caractéristiques qualitatives que quantitatives.

1.2.1 Distribution conjointe de deux caractéristiques

Il existe une façon de classer et ordonner deux caractéristiques à la fois pour en faire ressortir la distribution. Ainsi en combinant chaque catégorie d'une caractéristique à chaque catégorie de l'autre caractéristique et en trouvant l'effectif (ou la fréquence) de de ces combinaisons, nous arrivons à la distribution conjointe de deux caractéristiques à la fois.

1.2.2 Représentation de la distribution conjointe de deux caractéristiques

i) Tableau à double entrée

Une façon d'illustrer la distribution de deux caractéristiques à la fois, c'est d'employer un tableau à double entrée.

ii) Graphique

Pour représenter graphiquement cette distribution de deux caractéristiques à la fois, nous pouvons situer les catégories d'une caractéristique sur l'abcisse et les catégories de l'autre caractéristique sur l'ordonnée et par la suite, représenter chaque individu par un point. Cette représentation graphique s'appelle un nuage de points.

1.2.3 Recherche d'un lien linéaire entre deux caractéristiques quantitatives

Dans la recherche d'un lien il faut:

1) s'assurer de l'existence du lien,
2) déterminer son genre et sa force.

a) Outils pour évaluer le lien linéaire

i) outil visuel
- faire un nuage de points,
- vérifier si les points sont alignés.

ii) outil analytique
coefficient de corrélation (r)

$$r = \frac{\text{cov.}(x_i, y_i)}{(\text{é.-t. des } x_i)(\text{é.-t. des } y_i)}$$

caractéristiques du coefficient de corrélation

1. D'abord, il identifie l'existence d'un lien:

si $r = 0$ aucun lien possible,
si $r \neq 0$ un lien existe.

2. Il précise le genre de lien:

si $r < 0$ les valeurs des caractéristiques s'opposent,

si $r > 0$ les valeurs des caractéristiques concordent.

3. Il donne le degré de ce lien:

si $r \longrightarrow +1$ } c'est un lien
$r \longrightarrow -1$ } fort;

si $r \longrightarrow 0$ } c'est un lien faible.

Pour ce qui est de deux caractéristiques non quantitatives, on doit utiliser d'autres moyens pour mesurer le degré de dépendance entre ces caractéristiques.

b) Modèle pour exprimer le lien

Lorsqu'un lien linéaire existe entre deux caractéristiques quantitatives, on peut l'exprimer par la droite de régression linéaire. C'est la droite la plus proche possible (par le principe des moindres carrés) de l'ensemble des points du nuage.

On obtient une équation de la forme:

$$Y = mX + b$$

où $$m = \frac{\text{cov.}(x_i, y_i)}{\text{var. des } x_i}$$

$$b = \text{m.a.des } y_i - m(\text{m.a. des } x_i)$$

Résumé

Analyse de deux caractéristiques à la fois

1. Représentation de la distribution conjointe de deux caractéristiques

   a) Tableau à double entrée

   b) Graphique: nuage de points

2. Lien linéaire entre deux caractéristiques

   - Outils pour évaluer le lien linéaire

     a) Visuel - faire un nuage de points et vérifier si les points sont alignés.

     b) Analytique (coefficient de corrélation) (r)

$$r = \frac{\text{cov.}(x_i, y_i)}{(\text{é.-t. des } x_i)(\text{é.-t. des } y_i)}$$

$$-1 \leq r \leq 1$$

   - Modèle pour exprimer le lien linéaire

     Droite de régression linéaire

$$Y = mX + b$$

$$\text{où } m = \frac{\text{cov.}(x_i, y_i)}{\text{var. des } x_i}$$

$$b = \text{m.a. des } y_i - m(\text{m.a. des } x_i)$$

2. Analyse de l'information recueillie auprès d'une population

Si l'étude est faite à partir de tous les éléments de la population, on peut immédiatement passer à la rédaction du rapport, si au contraire l'étude ne peut être faite qu'à partir des éléments d'un échantillon, on doit relativer cette information à l'aide de la théorie d'échantillonnage.

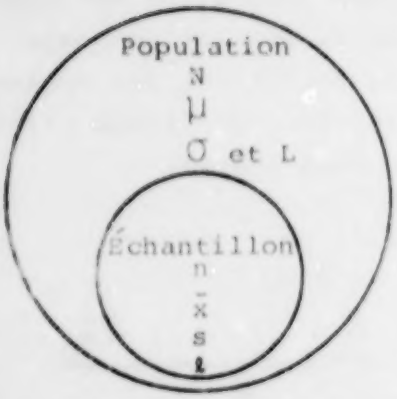
3. Analyse de l'information recueillie auprès d'un échantillon

La théorie d'échantillonnage est née du fait que dans de nombreuses circonstances, il est difficile sinon impossible de travailler avec tous les éléments de la population même après l'avènement des ordinateurs.

Les difficultés rencontrées sont principalement d'ordre économique, temporel et matériel. L'étude de la population est soit trop dispendieuse, trop longue ou impossible matériellement à réaliser en prenant tous les éléments; il s'agit de penser au recensement des morues dans l'estuaire du Saint-Laurent pour s'en convaincre. Que peut-on faire dans les circonstances? On pense naturellement à choisir un échantillon; ce dernier doit correspondre à nos attentes c'est-à-dire nous renseigner sur la population et rencontrer nos principales contraintes. Le choix de l'échantillon se joue entre la précision qu'on exige de l'échantillon et le degré de confiance qu'on veut avoir dans nos résultats.

Dans ce texte, on rappelle d'une façon succincte, rigoureuse et précise l'information dégagée aux laboratoires expérimental et d'applications concernant tout ce qui touche la théorie d'échantillonnage.

3.1 Les relations importantes entre l'échantillon et la population



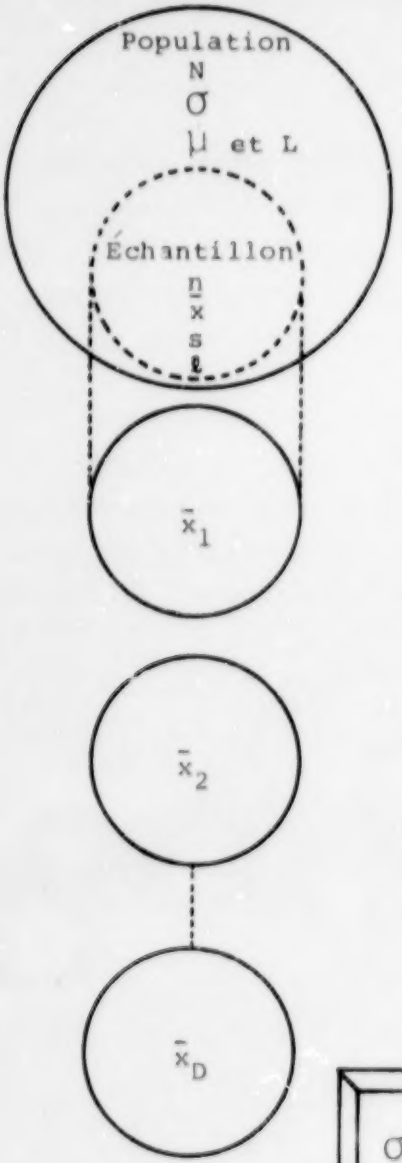
R 1 - La structure d'une ou de plusieurs caractéristiques de l'échantillon (s) n'est pas toujours identique à celle de la population (L).

R 2 - La tendance centrale et la dispersion d'une ou de plusieurs caractéristiques de l'échantillon ne sont pas toujours identiques à celles de la population.

Exemples:

1) $\bar{x} = \mu$
2) $s = \sigma$

} (pas toujours)



R 3 - Pour une caractéristique, la tendance centrale des moyennes arithmétiques des échantillons (taille n) possibles d'une population est égale à celle de la population quels que soient la taille (n) des échantillons et le genre de tirage effectué (avec ou sans remise)

$$\mu_{\bar{x}} = \mu$$

où

$$\mu_{\bar{x}} = \frac{\bar{x}_1 - \bar{x}_2 - \ldots - \bar{x}_D}{D}$$

R 4 - Pour une caractéristique, la dispersion des moyennes arithmétiques des échantillons possibles d'une population varie

- selon la taille des échantillons
- selon le mode de tirage

	Tirage	
	avec remise	sans remise
$\sigma_{\bar{x}} = ?$	$\frac{\sigma}{\sqrt{n}}$	$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

où

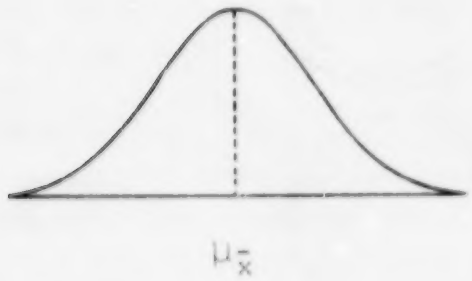
$$\sigma_{\bar{x}} = \sqrt{\frac{(\bar{x}_1 - \mu_{\bar{x}})^2 + (\bar{x}_2 - \mu_{\bar{x}})^2 + \ldots + (\bar{x}_D - \mu_{\bar{x}})^2}{D}}$$

R 5 - La distribution des moyennes arithmétiques des échantillons possibles d'une population varie

- selon la taille des échantillons,
- selon le mode de tirage,
- selon la structure de la population.

1) Si la taille de l'échantillon est élevée ($n \geq 30$), les moyennes arithmétiques ($\bar{x}$) des échantillons possibles sont distribuées selon une loi normale

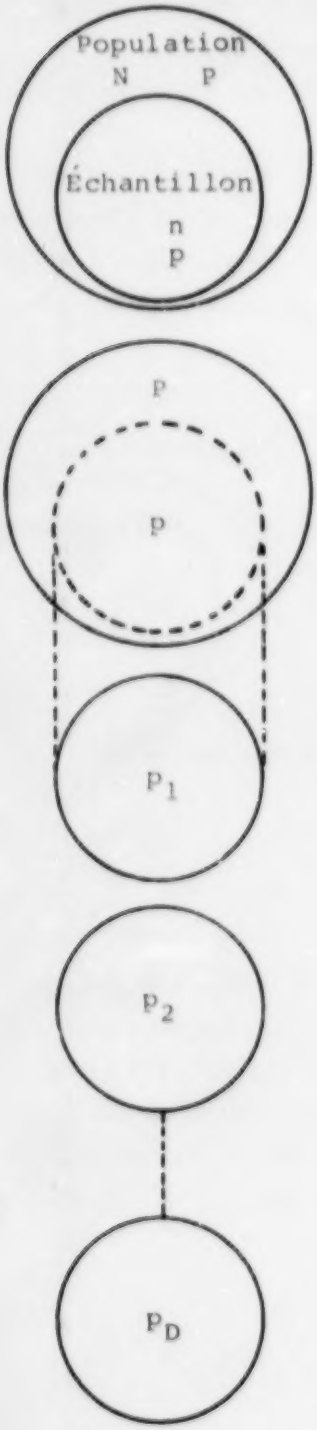
$$\bar{x} \sim \mathcal{N}(\mu_{\bar{x}}; \sigma_{\bar{x}})$$



$\mu_{\bar{x}}$

2) Si la taille de l'échantillon est petite ($n < 30$), les moyennes arithmétiques ($\bar{x}$) des échantillons possibles sont distribuées selon une loi normale seulement si la population est normale et la variance connue.

Analogiquement, tout ce qui est vrai pour les moyennes arithmétiques des échantillons l'est également pour les proportions des échantillons.



R 6 - La proportion d'une ou de plusieurs caractéristiques de l'échantillon n'est pas toujours identique à celle de la population.

$$p = P \quad \text{(pas toujours)}$$

R 7 - Pour une caractéristique, la tendance centrale des proportions des échantillons possibles d'une population est égale à celle de la population quels que soient la taille (n) des échantillons et le tirage effectué avec ou sans remise).

$$\mu_p = P$$

où $\mu_p = \dfrac{p_1 + p_2 + \ldots + p_D}{D}$

R 8 - Pour une caractéristique, la dispersion des proportions des échantillons possibles d'une population varie

- selon la taille des échantillons,
- selon le mode de tirage.

	Tirage	
	avec remise	sans remise
$\sigma_p = ?$	$\sqrt{\frac{PQ}{n}}$	$\sqrt{\frac{PQ}{n}}\ \sqrt{\frac{N-n}{N-1}}$

où

$$\sigma_p = \sqrt{\frac{(p_1 - \mu_p)^2 + (p_2 - \mu_p)^2 + \ldots + (p_D - \mu_p)^2}{D}}$$

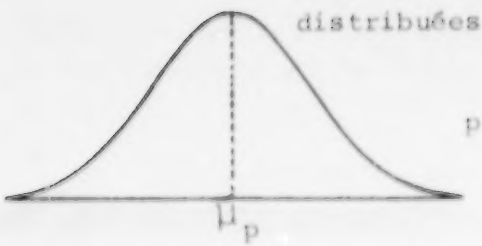
R 9 - La distribution des proportions des échantillons possibles d'une population varie

- selon la taille des échantillons,
- selon le mode de tirage.

1) Les proportions d'une caractéristique des échantillons possibles d'une population suivent la loi binômiale.

$$p \sim B(n; \mu_p)$$

2) Cependant, si nP et nQ sont grands (c'est-à-dire > 10, à la rigueur > 5), les proportions (p) d'une caractéristique des échantillons possibles sont distribuées selon une loi normale.



$$p \sim \mathcal{N}(\mu_p; \sigma_p)$$

La connaissance des précédentes relations nous permet de contrôler notre erreur d'échantillonnage et d'avoir une information valable sur la population étant donné nos contraintes.

Le théorème de limite centrale résume cette information.

Théorème de limite centrale

Considérons une population de moyenne μ (ou P) et de variance σ^2 (ou PQ). Si $\bar{X}$ désigne la moyenne d'un échantillon aléatoire de taille n, alors lorsque n augmente indéfiniment, $\bar{X}$ suit approximativement une loi normale de moyenne μ (ou P) et de variance σ^2/n (ou PQ/n) pourvu que la population soit infinie ou qu'elle soit finie et l'échantillon soit tiré avec remise.

Si la population est finie et l'échantillon tiré sans remise, la variance de $\bar{X}$ égale $\frac{\sigma^2}{n}\sqrt{\frac{N-n}{N-1}}$ où N est la taille de la population.

Résumé

Relations importantes entre l'échantillon et la population

1. La structure d'une caractéristique de l'échantillon peut différer de celle de la population.

$$\ell = L \text{ ou } \ell \neq L$$

2. La tendance centrale et la dispersion d'une caractéristique de l'échantillon peuvent différer de celles de la population.

$$\bar{x} = \mu \quad \text{ou} \quad \bar{x} \neq \mu$$
$$s = \sigma \quad \text{ou} \quad s \neq \sigma$$

3. La tendance centrale des moyennes arithmétiques des échantillons (taille n) d'une population est égale à celle de la population.

$$\mu_{\bar{x}} = \mu$$

4. La dispersion des moyennes arithmétiques des échantillons varie selon la taille et le mode de tirage des échantillons.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{(tirage avec remise)}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} \qquad \text{(tirage sans remise)}$$

5. La distribution des moyennes arithmétiques des échantillons d'une population est une loi normale.

   - si $n \geq 30$
   - si la population est normale.

6. La proportion d'une caractéristique de l'échantillon peut différer de celle de la population.

$$p = P \quad \text{ou} \quad p \neq P$$

7. La tendance centrale des proportions des échantillons d'une population est égale à celle de la population.

$$\mu_p = P$$

8. La dispersion des proportions des échantillons d'une population varie selon la taille et le mode de tirage des échantillons.

$$\sigma_p = \sqrt{\frac{PQ}{n}} \qquad \text{(tirage avec remise)}$$

$$\sigma_p = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N - n}{N - 1}} \qquad \text{(tirage sans remise)}$$

9. La distribution des proportions des échantillons d'une population suit

   - une loi binômiale
   - une loi normale, si $nP \cdot 5$ et $nQ \cdot 5$

3.2 Le langage pour exprimer l'erreur d'échantillonnage: la probabilité

L'étude porte toujours sur une ou des caractéristiques de la population; notre intérêt peut cependant se limiter à certaines valeurs de cette caractéristique, à un événement en particulier.

3.2.1 L'étude de l'événement

a) La probabilité d'un événement

La réalisation d'un événement et sa probabilité peuvent se résumer facilement dans un tableau comme on l'a vu à la symbolisation sur la probabilité.

Événement			
Réalisation	Impossible	Possible	
		Incertaine	Certaine
Probabilité	0	0 P(E) 1	1

b) Les méthodes pour évaluer la probabilité d'un événement (E)

1) Méthode directe

Un moyen élémentaire d'évaluer la chance de réaliser un événement est:

i) - d'examiner tout ce qui peut arriver dans l'expérience et de l'énumérer si possible,

$r_1, r_2, \ldots, r_N$ (les résultats possibles)

$S = \{r_1, r_2, \ldots, r_N\}$ (l'ensemble des résultats possibles)

ii) - d'identifier les résultats réalisant l'événement,

$E = \{x_1, x_2, \ldots, x_n\}$ (l'ensemble des résultats réalisant l'événement)

$E \subseteq S$

iii) - calculer la probabilité de chaque résultat réalisant l'événement,

$Pr(x_i) = ?$ $\quad 1 \leq i \leq n$

iv) - calculer la probabilité de l'événement en additionnant la probabilité de tous ceux qui le réalisent.

$Pr(E) = Pr(x_1) + Pr(x_2) + \ldots + Pr(x_n)$ où $x_i \in E$

Remarque:

Si $Pr(r_i) = Pr(r_j)$ et $i \neq j$,

alors $Pr(E) = \frac{\#E}{\#S}$

(On est alors dans un espace équiprobable.)

2) Méthodes indirectes

Les méthodes indirectes consistent à prendre un moyen détourné pour obtenir la probabilité de l'événement concerné.

i) $Pr(E) = 1 - Pr(\bar{E})$

ii) $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \cap E_2)$

iii) $Pr(E_1 \cap E_2) = Pr(E_1) Pr(E_2/E_1)$

Remarques

Ces méthodes se simplifient

1) si les deux événements sont incompatibles (ils ne peuvent pas être réalisés en même temps, $Pr(E_1 \cap E_2) = 0$),
alors

$Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2)$;

2) si les deux événements sont indépendants (la chance de réalisation de l'un n'est pas affectée par la réalisation de l'autre, $Pr(E_1/E_2) = Pr(E_1)$),
alors

$Pr(E_1 \cap E_2) = Pr(E_1) Pr(E_2)$.

3.2.2 L'étude de la caractéristique

On cherche alors la structure probabiliste de la caractéristique c'est-à-dire la probabilité de chacune de ses valeurs ou un moyen de l'obtenir, sa loi de probabilité, et les valeurs qui résument cette loi.

Dans l'espace probabiliste, des outils analogues à ceux utilisés dans l'espace fréquentiste existent; la probabilité de la valeur remplace la fréquence de la valeur.

Espace fréquentiste	Espace probabiliste
X: caractéristique (variable observée)	Y: caractéristique (variable aléatoire)
<u>Structure</u> de X	<u>Structure</u> de Y
1) Distribution des valeurs de X	1) Distribution des valeurs de Y
$\begin{array}{c\|c} X & x_1 \ \ldots \ x_n \\ \hline \text{fréquence} & f_1 \ \ldots \ f_n \end{array}$	$\begin{array}{c\|c} Y & y_1 \ \ldots \ y_n \\ \hline \text{probabilité} & Pr(y_1) \ldots Pr(y_n) \end{array}$
2) Résumé des valeurs de la caractéristique	2) Résumé des valeurs de la caractéristique
a) Tendance centrale	a) Tendance centrale
$m.a. = \sum_{i=1}^{n} f_i x_i$	$E(Y) = \sum_{i=1}^{n} Pr(y_i) y_i$
b) Dispersion	b) Dispersion
$\text{Variance} = \sum_{i=1}^{n} f_i (x_i - m.a.)^2$	$Var(Y) = \sum_{i=1}^{n} Pr(y_i) (y_i - E(Y))^2$

Exemple: la moyenne d'échantillon ($\bar{x}$) est une caractéristique de ce type; elle prend différentes valeurs dont la réalisation est due au hasard. On connaît sa moyenne et sa variance.

$$\mu_{\bar{x}} = E(\bar{x}) = \mu$$

$$\sigma^2_{\bar{x}} = \mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n} \qquad \text{(tirage avec remise)}$$

Résumé

La probabilité

1. Événement (E)

Réalisation	Impossible	Possible	
		Incertain	Certain
Probabilité	0	$0 < Pr(E) < 1$	1

si $E = \{x_1, x_2, \ldots, x_n\}$

alors $Pr(E) = Pr(x_1) + Pr(x_2) + \ldots + Pr(x_n)$

ou $x_i \in E$

De plus

a) $Pr(E) = 1 - Pr(\bar{E})$

b) $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \cap E_2)$

c) $Pr(E_1 \cap E_2) = Pr(E_1)\, Pr(E_2/E_1)$

2. Caractéristiques

Distribution des valeurs de Y

Y	Y_1	...	Y_n
probabilité	$Pr(Y_1)$	...	$Pr(Y_n)$

- $E(Y) = \sum_{i=1}^{n} Pr(Y_i) Y_i$

- $Var.(Y) = \sum_{i=1}^{n} Pr(Y_i)(Y_i - E(Y))^2$

3.3 L'analyse des erreurs possibles

L'analyse des erreurs repose essentiellement sur la connaissance de ce qui peut arriver lorsqu'on tire un échantillon d'une population, autrement dit de la connaissance des valeurs possibles c'est-à-dire de la distribution d'échantillonnage de la statistique considérée (moyenne ($\bar{x}$) ou proportion (p) d'échantillon).

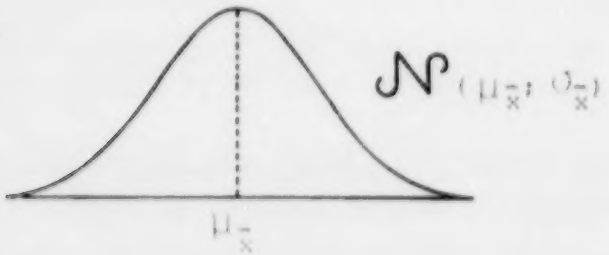
Puisque la valeur de l'échantillon est une des valeurs possibles, on n'est jamais certain que cette valeur corresponde à celle de la population.

On appellera, erreur d'échantillonnage, la différence entre la valeur obtenue de l'échantillon et la valeur réelle dans la population. Par exemple, la différence $\bar{x} - \mu$ est l'erreur d'échantillonnage de la moyenne arithmétique d'un échantillon par rapport au paramètre, la moyenne arithmétique de la population.

3.3.1 Les valeurs possibles de la moyenne d'échantillon:

La distribution d'échantillonnage des moyennes ($\bar{x}$)

Les différentes valeurs possibles de la moyenne d'échantillon sont connues (R 5); elles sont distribuées selon la loi normale centrée autour de la moyenne de la population (μ) (puisque $\mu_{\bar{x}} = \mu$) avec une dispersion $\sigma_{\bar{x}}$.



- si la taille de l'échantillon est élevée ($n \geq 30$);

- si la taille de l'échantillon est petite ($n < 30$),

  la population doit être normale et de variance connue.

Ainsi, la probabilité que l'erreur d'échantillonnage soit:

- plus élevée que 4 fois l'écart-type des $\bar{x}$ est presque nulle (événement presqu'impossible);

$$Pr(|\bar{x} - \mu| > 4\sigma_{\bar{x}}) \approx 0$$

- inférieure à 4 fois l'écart-type des $\bar{x}$ est pratiquement égale à un (événement certain):

$$Pr(|\bar{x} - \mu| < 4(\sigma_{\bar{x}})) = 1$$

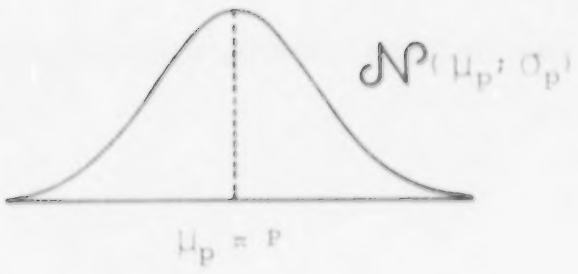
- inférieure à 1 fois l'écart-type des $\bar{x}$ est 0,6826 (événement possible).

$$Pr(|\bar{x} - \mu| < (\sigma_{\bar{x}})) = 0,6826$$

etc.

3.3.2 Les valeurs possibles de la proportion d'échantillon: la distribution d'échantillonnage des proportions (p)

Les différentes valeurs possibles de la proportion d'échantillon sont connues (R 9); elles sont distribuées selon la loi binômiale de moyenne (μ_p) et d'écart-type (σ_p). Cependant, si nP et nQ sont suffisamment grands, les proportions possibles des échantillons sont distribuées selon la loi normale.



Ainsi, la probabilité que l'erreur d'échantillonnage soit:

- plus élevée que 4 fois l'écart-type des p est presque nulle (événement presqu'impossible):

$$Pr(|p - P| \quad 4\sigma_p) \quad 0$$

- inférieure à 4 fois l'écart-type des p est pratiquement égale à un (événement certain):

$$Pr(|p - P| \quad 4\sigma_p) \quad 1$$

- inférieure à 1 fois l'écart-type des p est égale à 0,6826 (événement possible).

$$Pr(|p - P| \quad \sigma_p) = 0,6826$$

etc.

Résumé

Erreurs possibles en échantillonnage

1. Probabilité de l'erreur d'échantillonnage des moyennes

$$Pr(|\bar{x} - \mu| \leq 4(\sigma_{\bar{x}})) = 0$$

$$Pr(|\bar{x} - \mu| \leq 4(\sigma_{\bar{x}})) = 1$$

$$Pr(|\bar{x} - \mu| \leq (\sigma_{\bar{x}})) = 0,6826$$

2. Probabilité de l'erreur d'échantillonnage des proportions

$$Pr(|p - P| \leq 4(\sigma_p)) = 0$$

$$Pr(|p - P| \leq 4(\sigma_p)) = 1$$

$$Pr(|p - P| \leq (\sigma_p)) = 0,6826$$

3.4 Le choix d'un échantillon pour représenter correctement la population selon l'erreur acceptable et le degré de confiance désiré.

Un échantillon est valable s'il est représentatif de la population et suffisamment significatif. Pour l'obtenir il faut répondre à cette question:

1) Qui choisir?
2) Comment les choisir?
3) Combien en choisir?

3.4.1 Les questions

- *Qui choisir?*

Tout élément de la population: chacun doit avoir une chance calculable de faire partie de l'échantillon.

1) Cette chance n'a pas besoin d'être calculée. Il suffit qu'elle soit calculable.

2) Cela n'implique pas que chaque individu ait une cha..ce égale. C'est le cas cependant de la méthode aléatoire simple. Mais parfois la recherche de l'efficacité et les besoins de l'étude conduisent à abandonner ce type d'échantillonnage.

- *Comment les choisir?*

Selon l'une des méthodes suivantes:

- l'échantillonnage simple aléatoire,
- l'échantillonnage systématique,
- l'échantillonnage stratifié aléatoire,
- l'échantillonnage par grappes,
- l'échantillonnage à 2, 3, . ., n degrés.

- *Combien en choisir?*

La quantité nécessaire exigée par la précision et le degré de confiance désirés.

3.4.2 Les étapes

On doit

1) choisir la méthode d'échantillonnage;

- Le choix se fait d'après les besoins et le contexte de l'étude.

2) établir la base de sondage;

- La liste des éléments de la population nécessaire à la construction de l'échantillon.

3) déterminer la quantité d'éléments de l'échantillon;

- L'emploi des relations suivantes:

a) pour la moyenne d'une caractéristique de la population.

$$n = \begin{cases} z_c^2 \dfrac{\sigma^2}{V^2} & \text{(tirage avec remise)} \\ \dfrac{z_c^2 \, N \, \sigma^2}{(N-1)V^2 + z_c^2 \sigma^2} & \text{(tirage sans remise)} \end{cases}$$

n: taille de l'échantillon
N: taille de la population
σ^2: variance de la population
z_c: valeur de la variable z qui détermine le degré de confiance d'une
V: valeur de la précision ou de l'erreur acceptable

Preuve:

La précision = V = $z_c \, \sigma_{\bar{x}}$
(ou l'erreur)

i) tirage avec remise

$$V = z_c \, \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

$$V\sqrt{n} = z_c \sigma$$

$$\sqrt{n} = z_c \frac{\sigma}{V}$$

d'où $n = z_c^2 \dfrac{\sigma 2}{V^2}$

ii) tirage sans remise

$$V = z_c \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

$$V^2 = z_c^2 \frac{\sigma^2}{n} \left(\frac{N - n}{N - 1} \right) = \frac{z_c^2 \sigma^2}{(N-1)} \left(\frac{(N-n)}{n} \right)$$

$$V^2 = \frac{z_c^2 \sigma^2}{(N-1)} \left(\frac{N}{n} - 1 \right) = \frac{z_c^2 \sigma^2 N}{(N-1) n} - \frac{z_c^2 \sigma^2}{(N-1)}$$

$$V^2 + \frac{z_c^2 \sigma^2}{(N-1)} = \frac{z_c^2 \sigma^2 N}{(N-1) n}$$

$$n = \frac{z_c^2 N \sigma^2}{(N-1) \left[V^2 + \frac{z_c^2 \sigma^2}{(N-1)} \right]}$$

$$n = \frac{z_c^2 N \sigma^2}{(N-1) \left[\frac{(N-1) V^2 + z_c^2 \sigma^2}{(N-1)} \right]}$$

d'où $$n = \frac{z_c^2 N \sigma^2}{(N-1) V^2 + z_c^2 \sigma^2}$$

Remarque

Si la variance de la population (σ^2) est inconnue, on l'estime à partir de la variance d'un échantillon (s^2) d'une taille arbitraire d'au moins 30 éléments.

b) *pour la proportion d'une caractéristique de la population*

$$n = \begin{cases} \dfrac{z_\alpha^2 \; P \; Q}{V^2} & \text{(tirage avec remise)} \\[2ex] \dfrac{z_\alpha^2 \; N \; P \; Q}{(N-1)V^2 + z_\alpha^2 \; P \; Q} & \text{(tirage sans remise)} \end{cases}$$

n: taille de l'échantillon
N: taille de la population
P: proportion de la catégorie étudiée de la population
Q: 1 - P
z_α: *valeur de la variable z qui détermine le degré de confiance désiré*
V: valeur de la précision ou de l'erreur acceptable

Remarque:

Si P est inconnu, on l'estime par $\frac{1}{2}$ puisque la taille de l'échantillon (n) est maximum lorsque $P = \frac{1}{2}$. On est ainsi assuré d'avoir la précision et le degré de confiance désirés.

4) Construire l'échantillon.

On se sert d'un processus qui conduit à l'obtention d'observations indépendantes de la loi discrète uniforme, celle donnant à chacun des dix chiffres: 0, 1, ..., 9 la probabilité d'apparition de 1/10.

L'utilisation d'une table de nombres au hasard est une façon très simple de construire l'échantillon.

En somme, les éléments de l'échantillon sont choisis au hasard (qui?) selon la méthode adaptée aux besoins de l'étude (comment?) pour répondre à la précision et au degré de confiance désirés (combien?).

Résumé

Choix d'un échantillon

1. Qui choisir?

   Chaque élément doit avoir une chance calculable de faire partie de l'échantillon.

2. Comment les choisir?

   Selon la méthode de l'échantillonnage simple aléatoire.

3. Combien en choisir?

   n = ? a) Pour la moyenne

$$= \frac{z_c^2 \, \sigma^2}{V^2} \qquad \text{(tirage avec remise)}$$

$$= \frac{z_c^2 \, N\sigma^2}{(N-1)V^2 + z_c^2 \, \sigma^2} \qquad \text{(tirage sans remise)}$$

   b) Pour la proportion

$$= \frac{z_c^2 \, PQ}{V^2} \qquad \text{(tirage avec remise)}$$

$$= \frac{z_c^2 \, NPQ}{(N-1)V^2 - z_c^2 \, PQ} \qquad \text{(tirage sans remise)}$$

3.5 L'analyse de l'information recueillie dans l'échantillon

L'étude de l'information recueillie dans l'échantillon suit la même démarche que celle tirée d'une population. On procède en étudiant une caractéristique à la fois, ou deux, ..., à la fois.

On peut

1) trouver la distribution de chaque caractéristique,

2) résumer les informations obtenues de l'échantillon sur la(les) caractéristique(s) par quelques valeurs,

3) faire entrer toutes les informations dans une formule mathématique ou encore trouver un modèle qui représente valablement les observations.

C'est un moyen puissant qui permet:

- de manipuler facilement un ensemble complexe,
- de rechercher des optimums,
- d'intrapoler, etc.

De plus, si cette formule est simple, elle peut ne dépendre que d'un nombre réduit de paramètres. Donc, pour la connaître, il suffit de pouvoir estimer ces paramètres.

3.6 L'inférence de l'information obtenue de l'échantillon à la population

Le problème est de tirer des résultats de l'analyse de l'échantillon, des informations valables pour l'ensemble de la population.

Ces résultats permettent rarement d'être sûr des propriétés de la population; elle ne donne que des indices sur la population. Il s'agit d'évaluer jusqu'à quel point ces indices sont valables. En général la valeur des indices s'exprime en termes de probabilité.

"Plus la taille de l'échantillon est grande, plus la différence entre l'échantillon et la population a de chances d'être petite."

Ayant tiré les éléments de l'échantillon au hasard, on peut considérer l'échantillon comme une réalisation d'une épreuve aléatoire qui obéit à la loi des grands nombres.

Le théorème de Tchébicheff dit:

$$Pr\left[\,\left|k/n - \bar{w}\right| \cdot \varepsilon\,\right] \cdot \frac{1}{4}\, n\, \varepsilon^{2}$$

Il montre qu'au fur et à mesure que n augmente, on est de plus en plus certain que la différence entre le rapport k/n et la tendance centrale $\bar{w}$ est petite. Cependant, elle n'est pas nécessairement égale à la tendance centrale de la population (P). Il faut alors que le hasard soit réalisé d'une manière parfaite. Ces conditions préalables déterminent la valeur de l'inférence.

Il existe deux situations: l'une où on n'a aucune connaissance a priori de la population et l'autre où on a une connaissance a priori de certaine éléments de la population.

3.6.1 Aucune connaissance a priori de la population

On veut connaître la population, en particulier la valeur de ses paramètres. On en fait une estimation.

Les genres d'estimation

a) l'estimation ponctuelle

On se sert d'une valeur unique qui doit se rapprocher le plus du paramètre à estimer.

Paramètre	Estimateur
μ	$\bar{x}$
P	p
σ^2	s^2

Un bon estimateur doit être entre autres sans biais et efficace.

$\bar{x}$ et p : deux bons estimateurs
s^2 doit être corrigé par $\frac{n}{n-1}$. Ainsi $s^2\left(\frac{n}{n-1}\right)$ est un bon estimateur de σ^2.

La mesure de l'erreur dans ce genre d'estimation est impossible.

b) l'estimation par intervalle de confiance

L'information fournie par l'échantillon sert d'indice, de base, pour la construction de l'intervalle de confiance; on ajoute et on soustrait une valeur, selon le degré de confiance désiré, à l'information dans l'échantillon.

1) $\mu = ?$

$\bar{x} \pm V$

où

$$V = z_c \sigma_{\bar{x}} \quad \text{et} \quad \sigma_{\bar{x}} = \begin{cases} \dfrac{\sigma}{\sqrt{n}} & \text{(tirage avec remise)} \\ \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}} & \text{(tirage sans remise)} \end{cases}$$

2) $P = ?$

$p \pm V$

où

$$V = z_c \sigma_p \quad \text{et} \quad \sigma_p = \begin{cases} \sqrt{\dfrac{P\,Q}{n}} & \text{(tirage avec remise)} \\ \sqrt{\dfrac{P\,Q}{n}} \sqrt{\dfrac{N-n}{N-1}} & \text{(tirage sans remise)} \end{cases}$$

3.6.2 Connaissance a priori de certains éléments de la population.

On a des connaissances a priori de la population et on veut vérifier la validité de ces connaissances à l'aide d'un échantillon.

On raisonne de la même façon pour les paramètres que pour la structure d'une ou des caractéristiques de la population.

a) Principe des tests

Deux hypothèses sont émises, l'hypothèse nulle et l'hypothèse alternative ou la contre-hypothèse.

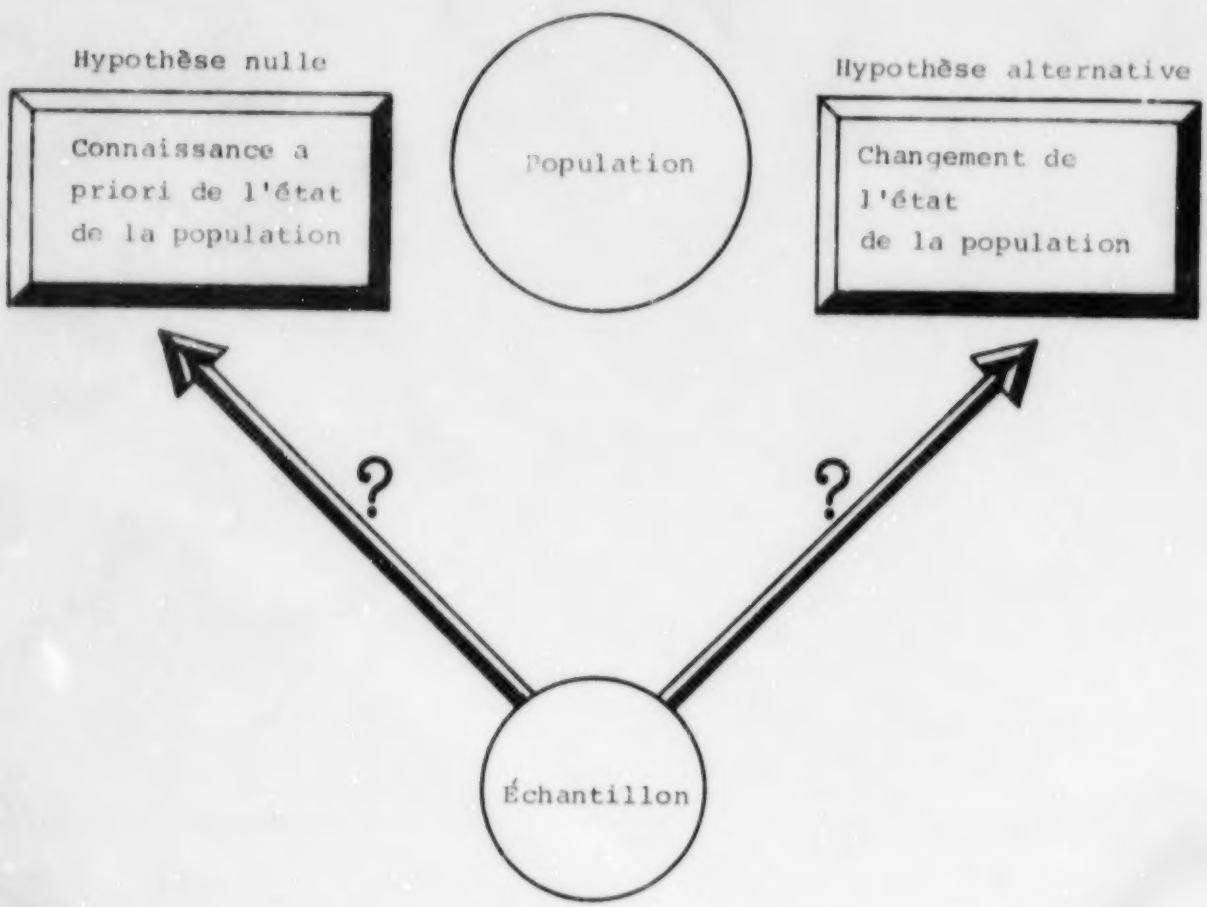
L'hypothèse nulle correspond habituellement à nos connaissances a priori de l'état de la population.

L'hypothèse alternative correspond à ce qu'on veut vérifier c'est-à-dire le changement d'état de la population.

On vérifie si la chance d'obtenir le genre d'information contenue dans l'échantillon est élevée ou faible quand l'hypothèse nulle est vraie.

- Si cette chance est élevée, il semble vraisemblable d'accepter l'hypothèse nulle.

- Par contre, si cette chance est faible, il semble vraisemblable de la rejeter et d'accepter l'hypothèse alternative.



b) Le procédé technique des tests d'hypothèses

1) On émet les hypothèses.

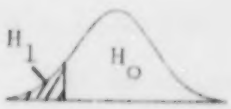
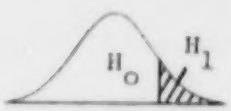
2) On fixe le niveau de signification du test (α).

3) On établit les règles de décision pour choisir les hypothèses (ou on fixe les régions d'acceptation et de rejet de l'hypothèse nulle).

4) On analyse l'information de l'échantillon.

5) On conclut selon les règles de décision établies auparavant.

c) Les genres d'hypothèses et les régions critiques

Paramètre	Genre d'hypothèses	Région critique	Test
μ ou p	1. H_o: par = V H_1: par ≠ V		Unilatéral
	2. H_o: par = V H_1: par ≠ V		
	3. H_o: par = V H_1: par ≠ V		Bilatéral

{ V: une valeur fixe
par: le paramètre

Résumé

Inférence de l'échantillon à la population

1. Aucune connaissance a priori de la population

   a) Estimation de μ

      - ponctuelle $\bar{x}$
      - par intervalle de confiance

        $\bar{x} \pm z_c \, \sigma_{\bar{x}}$

   b) Estimation de P

      - ponctuelle p
      - par intervalle de confiance

        $P \pm z_c \, \sigma_p$

2. Connaissance a priori de certains éléments de la population

   a) test d'hypothèses sur μ

   b) test d'hypothèses sur P

4. Rédaction d'un rapport

Après l'analyse de l'information recueillie, soit auprès d'une population, soit auprès d'un échantillon, il faut rédiger un rapport dont les étapes sont les suivantes:

- L'avant-propos ou la préface

- L'introduction

- Le corps du rapport
  - les données et la méthodologie
  - l'analyse statistique des données
  - la discussion des résultats

- La conclusion

- Les annexes

TABLE DES MATIERES

Introduction

1. Analyse d'un groupe de données F.9

1.1 Analyse d'une caractéristique à la fois: recherche de la structure de la caractéristique F.9

1.1.1 Nature d'une caractéristique F.9

1.1.2 Distribution d'une caractéristique F.11

1.1.3 Représentation de la distribution d'une caractéristique F.11
- Tableau à simple entrée F.11
- Graphiques F.15

1.1.4 Synthèse de l'information F.18
- Tendance centrale F.18
- Dispersion F.24

1.1.5 Modèles théoriques de l'information F.33

1.2 Analyse de deux caractéristiques à la fois: recherche de la structure conjointe des deux caractéristiques F.37

1.2.1 Distribution conjointe des deux caractéristiques F.37

1.2.2 Représentation de la distribution conjointe ..... F 37
- Tableau à double entrée F.37
- Graphique F.37

1.2.3 Recherche d'un lien linéaire entre deux caractéristiques quantitatives..... F.38
- Outil pour évaluer le lien linéaire..F.38
- Modèle pour exprimer le lien.....F.39

2. Analyse de l'information recueillie auprès d'une population F.41

3. Analyse de l'information recueillie auprès d'un échantillon F.42

3.1 Relations importantes entre l'échantillon et la population F.43

3.2 Langage pour exprimer l'erreur d'échantillonnage: la probabilité F.51

3.2.1 Étude de l'événement F.51
- Probabilité d'un événement F.51
- Méthodes pour évaluer la probabilité d'un événement (E)..... F.51

3.2.2. Étude de la caractéristique F.54

3.3 Analyse des erreurs possibles F.56

3.3.1 Les valeurs possibles de la moyenne d'échantillon: distribution d'échantillonnage des moyennes F.58

3.3.2 Les valeurs possibles de la proportion d'échantillon: distribution d'échantillonnage des proportions..... F.59

3.4 Le choix d'un échantillon pour représenter correctement la population selon l'erreur acceptable et le degré de confiance désiré ... F.62

3.4.1. Les questions F.62

3.4.2. Les étapes F.63

3.5 Analyse de l'information recueillie dans l'échantillon F.69

3.6 Inférence de l'information obtenue de l'échantillon à la population F.69

3.6.1 Aucune connaissance a priori de la population F.70
- Estimation ponctuelle F.71
- Estimation par intervalle de confiance..... F.71

3.6.2 Connaissance a priori de certains éléments de la population F.72
- Principe des tests d'hypothèses..... F.72
- Procédé technique des tests d'hypothèses F.74
- Genres d'hypothèses et les régions critiques F.75

4. Rédaction d'un rapport F.77

Annexes

Enquête

Activités estivales des étudiants(es) du groupe

QUESTIONNAIRE

1. IDENTIFICATION

1.1 Sexe
1.2 Niveau scolaire
1.3 Secteur
1.4 Orientation
1.5 Âge
1.6 Date de naissance
1.7 Lieu de résidence
1.8 Profession du père
1.9 Profession de la mère

2. TRAVAIL

2.1 As-tu travaillé durant tes dernières vacances estivales? Si non, passe à la partie ETUDE.
2.2 Comment as-tu trouvé ton emploi?
2.3 Quel était le type de ton travail?
2.4 Ton travail était-il connexe à tes études?
2.5 Dans quel milieu as-tu travaillé?
2.6 À combien de kilomètres de ta résidence permanente était situé ton lieu de travail?
2.7 Combien de semaines a duré ton travail?
2.8 Combien de jour(s) par semaine as-tu travaillé?
2.9 Combien d'heures par jour as-tu travaillé?
2.10 As-tu été payé pour ton travail?
2.11 Si oui, combien as-tu gagné en tout?

3. ÉTUDE

3.1 As-tu étudié durant tes dernières vacances estivales? Si non, passe à la partie VOYAGES.

3.2 À quelle institution as-tu étudié?

3.3 Pourquoi as-tu étudié durant tes vacances?

3.4 Pendant combien de semaines as-tu étudié?

3.5 Combien de jours par semaine as-tu étudié?

3.6 Combien d'heures par jour as-tu étudié?

3.7 As-tu suivi des cours du jour ou du soir?

3.8 Dans quelle(s) matière(s) as-tu étudié?

3.9 Quelles dépenses t'ont occasionné tes études?

4. VOYAGES

4.1 As-tu voyagé durant tes dernières vacances estivales? Si non, passe à AUTRES ACTIVITES.

4.2 Où es-tu allé en voyage?

4.3 Quelle distance en kilomètres as-tu parcourue dans ces voyages?

4.4 Quel moyen de transport as-tu utilisé?

4.5 Durant combien de jours as-tu voyagé?

4.6 Quelles dépenses t'ont occasionné tes voyages?

5. AUTRES ACTIVITÉS

5.1 Nom(s) d'autre(s) activité(s)?

5.2 Combien d'heures as-tu allouées à ces activités?

5.3 Quelles dépenses t'ont occasionné ces activités?

6. SATISFACTION

6.1 Es-tu content(e) de tes vacances?

QUESTIONNAIRE NUMÉRO: _____

RÉPONSES AU QUESTIONNAIRE:

1.1 Masculin ☐ Féminin ☐

1.2 Coll. I ☐ Coll. II ☐ Coll. III ☐

1.3 Général ☐ Professionnel ☐

1.4 Orientation ☐

1.5 Âge ☐☐ ans

1.6 ☐☐☐

1.7 Urbain (ville) ☐ Rural ☐

1.8 _____

1.9 _____

2.1 Oui ☐ Non ☐

2.2 Bureau de placement ☐ Même emploi que l'année précédente ☐
Journaux ☐ Amis ☐
Famille ☐
Autres ☐

2.3 Manuel ☐ Intellectuel ☐

2.4 Pas du tout ☐ Beaucoup ☐
Un peu ☐ Complètement ☐

2.5 Secteur public ☐ Famille ☐
Secteur privé ☐ Chez des amis ☐

2.6 Moins de 3 km ☐ 10 à 50 km ☐
3 à 10 km ☐ Plus de 50 km ☐

2.7 ☐ semaines

2.8 ☐ jours

2.9 ☐ heures

2.10 Oui ☐ Non ☐

2.11 $ ☐

3.1 Oui ☐ Non ☐

3.2 Institution ☐

3.3 _____

3.4 ☐ semaine(s)

3.5 ☐ jour(s)

3.6 ☐ heure(s)

3.7 Jour ☐ Soir ☐

3.8 _____

3.9 Coût total $ ☐

4.1 Oui ☐ Non ☐

4.2 Québec ☐ États Unis ☐
Autres provinces ☐ Ailleurs ☐

4.3 ☐ km

4.4 Avion ☐ Automobile ☐ Moto ☐
Train ☐ Autobus ☐ ☐
Bateau ☐ "Pouce" ☐

4.5 ☐ jours

4.6 Coût total $ ☐

5.1 _____

_____

_____

5.2 ☐ heures
☐ "
☐ "

5.3 $ ☐
$ ☐
$ ☐

6.1 Pas du tout ☐ Assez ☐
Très peu ☐ Beaucoup ☐

Manuel d'instructions

Table des matières

1. LES OUTILS STATISTIQUES

- Comment construire un tableau? M. 3
- Comment faire une représentation graphique? M. 9
- Comment mesurer autrement la dispersion? M. 20
- Comment représenter une distribution par un modèle théorique? M. 21
- Comment travailler avec une distribution normale? M. 28
- Comment trouver la structure conjointe de 2 caractéristiques? M. 43
- Comment choisir un échantillon d'une population? M. 51
- Comment estimer un paramètre d'une population? M. 69
- Comment tester une hypothèse statistique? M. 81

2. LA MÉTHODE POUR ÉTUDIER DES DONNÉES M. 97

3. LA RÉDACTION D'UN RAPPORT M. 103

1. Les outils statistiques

Comment construire un tableau?

L'utilité

Présenter d'une façon synthétique un grand nombre de données afin de les analyser plus facilement.

Les étapes de construction

E.1 Choisir une caractéristique.

E.2 Énumérer et ordonner, si possible, les catégories différentes.

E.3 Donner l'effectif de chaque classe.

E.4 Donner l'effectif total.

E.5 Écrire un titre, une date, et la provenance des données s'il y a lieu.

Tableau 1 La répartition de la population du Québec selon l'état matrimonial (1976)

ÉTAT MATRIMONIAL	EFFECTIF
CÉLIBATAIRES DE MOINS DE 15 ANS	1 550 335
CÉLIBATAIRES DE 15 ANS OU PLUS	1 451 220
MARIÉ(E)S	2 906 970
VEUFS - VEUVES	259 940
DIVORCÉ(E)S	65 970
TOTAL	6 234 435

SOURCE: Recensement du Canada, 1976

Afin de comparer deux groupes qui n'ont pas le même effectif total, il est utile d'exprimer les effectifs en fréquences (pourcentages).

Tableau 2 La population urbaine selon la langue maternelle Québec-1976

LANGUE MATERNELLE	EFFECTIF	FRÉQUENCE
FRANÇAIS	3 823 530	77,5%
ANGLAIS	701 530	14,2%
AUTRES	407 695	8,3%
TOTAL	4 932 755	100%

SOURCE: Recensement du Canada, 1976

Tableau 3 La population rurale selon la langue maternelle

Québec-1976

LANGUE MATERNELLE	EFFECTIF	FRÉQUENCE
FRANÇAIS	1 165 465	89,5%
ANGLAIS	99 150	7,6%
AUTRES	37 075	2,9%
TOTAL	1 301 690	100%

SOURCE: Recensement du Canada, 1976

Lorsque le nombre de catégories différentes est très élevé, il convient de regrouper les données en classes, chaque classe contenant un certain nombre de catégories différentes. La décision concernant le nombre de classes est surtout une question d'expérience et de jugement et dépend du problème traité.

Dans le cas des catégories numériques, les classes doivent, autant que possible, avoir la même largeur et être contiguës (juxtaposées). Le tableau 4 en est un exemple.

Tableau 4 La population selon l'âge
Québec-1976

ÂGE (ANS)	EFFECTIF	FRÉQUENCE (%)
0 - 4	442 655	7,10
5 - 9	485 610	7,79
10 - 14	622 065	9,98
15 - 19	666 290	10,69
20 - 24	598 440	9,60
25 - 29	555 255	8,91
30 - 34	474 555	7,61
35 - 39	374 810	6,01
40 - 44	354 085	5,68
45 - 49	349 315	5,60
50 - 54	327 340	5,25
55 - 59	268 790	4,31
60 - 64	233 880	3,75
65 - 69	186 070	2,98
70 - 74	133 005	2,13
75 - 79	85 610	1,37
80 - 84	46 850	0,75
85 - 89	21 110	0,34
90 et plus	8 715	0,14
TOTAL	6 234 450	99,99

SOURCE: Recensement du Canada 1976

La même distribution peut être représentée par un tableau des effectifs cumulés ou des fréquences cumulées

Tableau 5 La population selon l'âge

Québec - 1976

ÂGE (ANS)	EFFECTIF CUMULÉ	FRÉQUENCE CUMULÉE (%)
moins de 4,5	442 655	7,10
moins de 9,5	928 265	14,89
moins de 14,5	1 550 330	24,87
moins de 19,5	2 216 620	35,56
moins de 24,5	2 815 060	45,16
moins de 29,5	3 370 315	54,07
moins de 34,5	3 844 870	61,68
moins de 39,5	4 219 680	67,69
moins de 44,5	4 573 765	73,37
moins de 49,5	4 923 080	78,97
moins de 54,5	5 250 420	84,22
moins de 59,5	5 519 210	88,53
moins de 64,5	5 753 090	92,28
moins de 69,5	5 939 160	95,26
moins de 74,5	6 072 165	97,39
moins de 79,5	6 157 775	98,76
moins de 84,5	6 204 625	99,51
moins de 89,5	6 225 735	99,85
moins de 94,5	6 234 450	99,99

Ce tableau nous informe qu'en 1976, il y avait 2 216 620 Québécois âgés de moins de 19,5 ans ce qui représente 35,56% de la population totale au Québec.

Les valeurs 4,5; 9,5;...; 94,5 représentent la borne supérieure de chaque classe.

Comment faire une représentation graphique?

L'utilité

Présenter d'une façon visuelle un phénomène donné.

Les principes généraux de construction

Tout graphique doit

a) être esthétique, agréable à regarder,

b) bien représenter la réalité.

Les différentes représentations graphiques

1. Diagramme en bâtons

Les étapes de construction

E.1 *On trace un système d'axes (le premier quadrant seulement).*

E.2 *Sur l'abscisse, on indique les catégories différentes et au bout de l'axe, on indique la caractéristique.*

E.3 *On gradue l'axe vertical de sorte que la valeur maximale soit à peu près égale à la valeur de l'effectif partiel le plus élevé ou la fréquence la plus élevée. On écrit aussi le mot effectif ou fréquence ou ce qu'il représente.*

E.4 *Vis-à-vis de chaque catégorie, à partir de l'abcisse, on élève un bâton dont la hauteur est proportionnelle à l'effectif ou la fréquence.*

E.5 *On écrit un titre, une date et la source des données s'il y a lieu.*

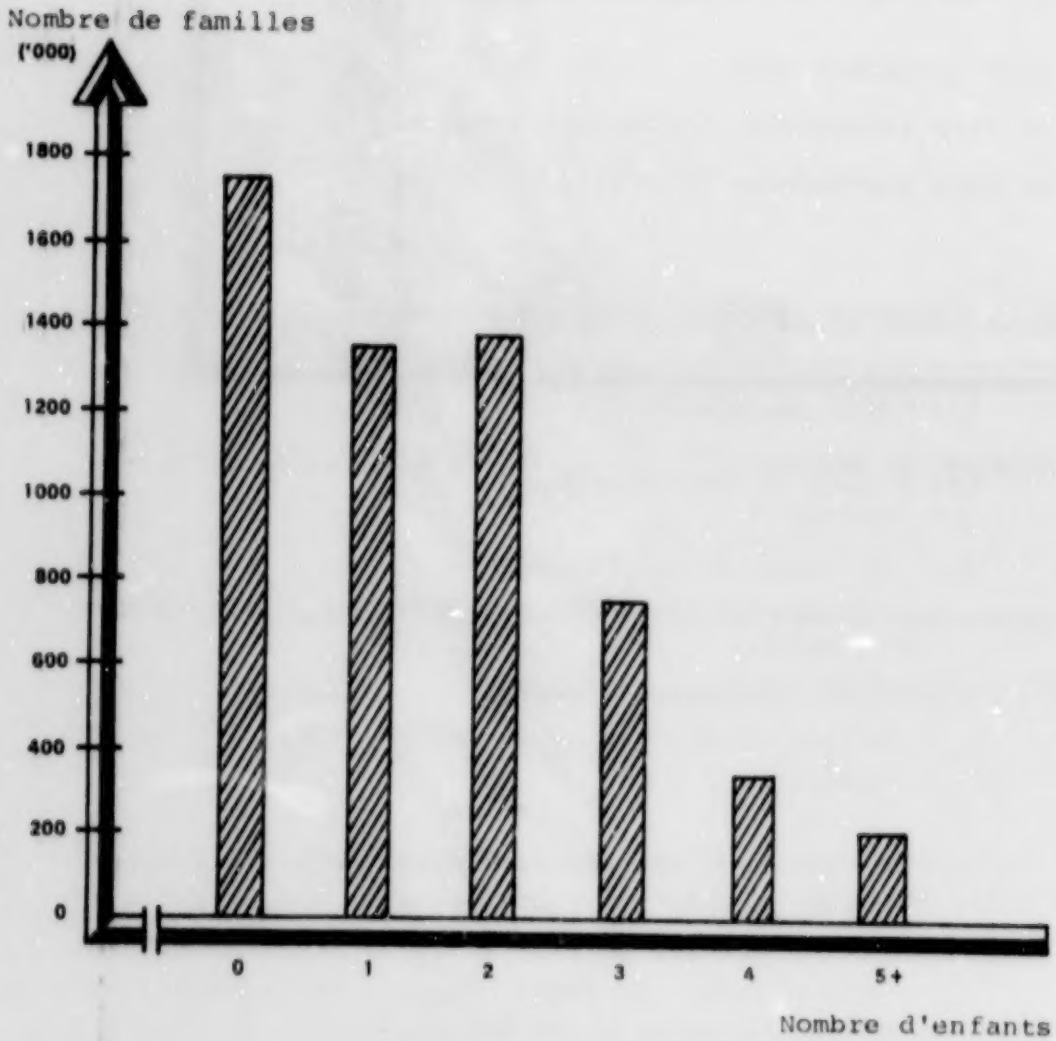


Figure 1 La répartition des familles selon le nombre d'enfants à la maison

Canada-1976

SOURCE: Recensement du Canada-1976

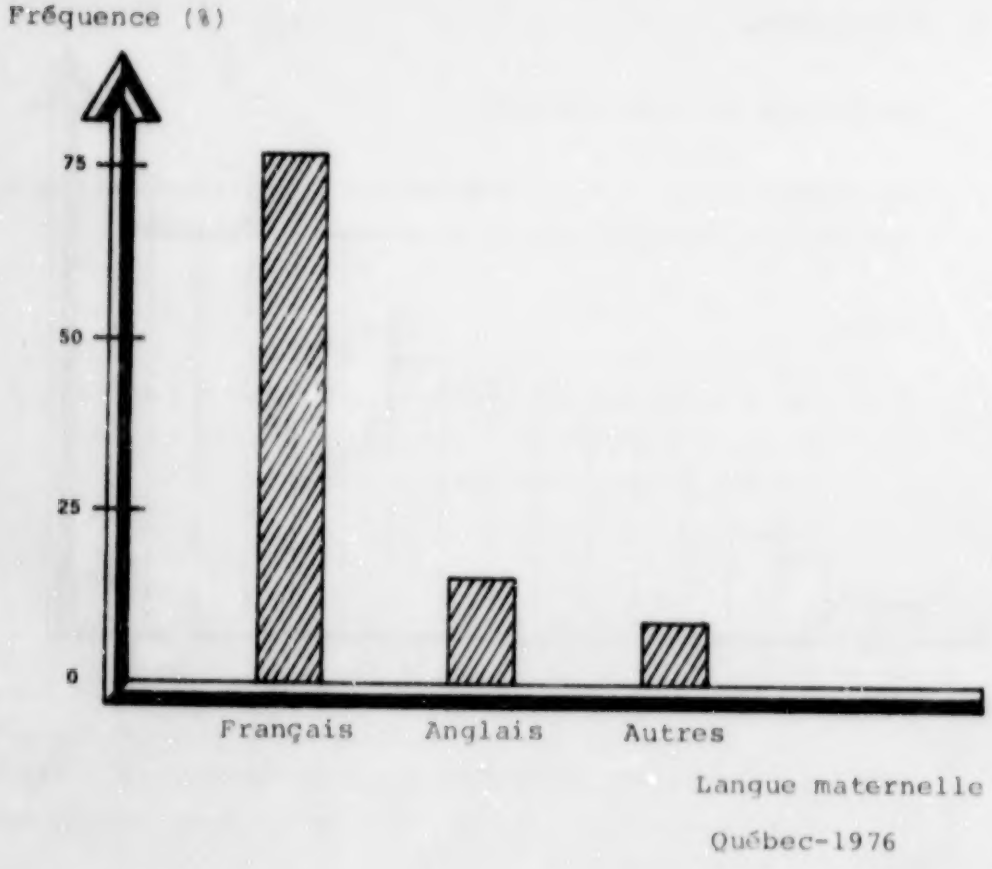


Figure 2 La population urbaine selon la langue maternelle
Québec 1976
SOURCE: Recensement du Canada - 1976

Le diagramme en bâtons se prête bien à la représentation d'une caractéristique dont le nombre de catégories différentes n'est pas trop élevé.

Le signe || sur l'abcisse de la figure 1 indique un déplacement de cet axe vers la droite.

On peut utiliser le même symbole pour représenter un déplacement de l'abcisse vers la gauche.

2. Histogramme

Les étapes de construction

Les étapes 1, 3, 5 sont semblables à celles décrites lors de la construction du diagramme en bâtons.

E.2 *Sur l'abcisse, on indique la valeur du milieu de chaque classe et au bout de l'axe, on indique la caractériatique.*

E.4 *Pour chaque classe, on élève un rectangle dont le centre de la base correspond au milieu de la classe. L'aire du rectangle doit être proportionnelle à l'effectif ou la fréquence de la classe. Enfin, les rectangles successifs doivent être juxtaposés.*

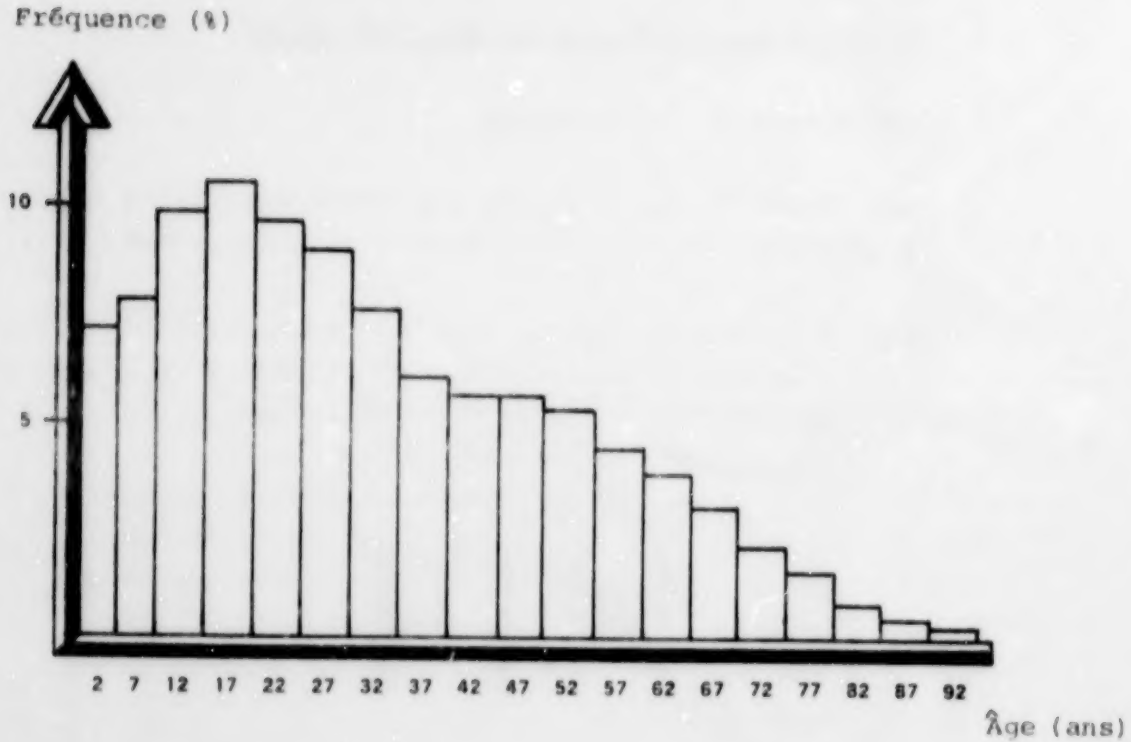


Figure 3 La population du Québec selon l'âge, 1976

SOURCE: Recensement du Canada - 1976

Remarque: la classe 90 et plus est représentée par la valeur 92 sur ce graphique.

L'histogramme se prête bien à la représentation graphique de données regroupées en classes.

3. Polygone des effectifs ou des fréquences

Les étapes de construction

Les étapes 1, 2, 3, 5 sont les mêmes que celles utilisées lors de la construction d'un histogramme.

E.4 *Pour chaque classe, on place un point dont l'abcisse vaut le milieu de la classe et l'ordonnée vaut l'effectif ou la fréquence. On relie, par des segments de droite, les points successifs. Pour fermer le polygone sur l'abcisse on trouve deux points: le milieu de la classe qui précède la classe inférieure et le milieu de la classe qui suit la classe supérieure. L'effectif de ces deux classes est zéro.*

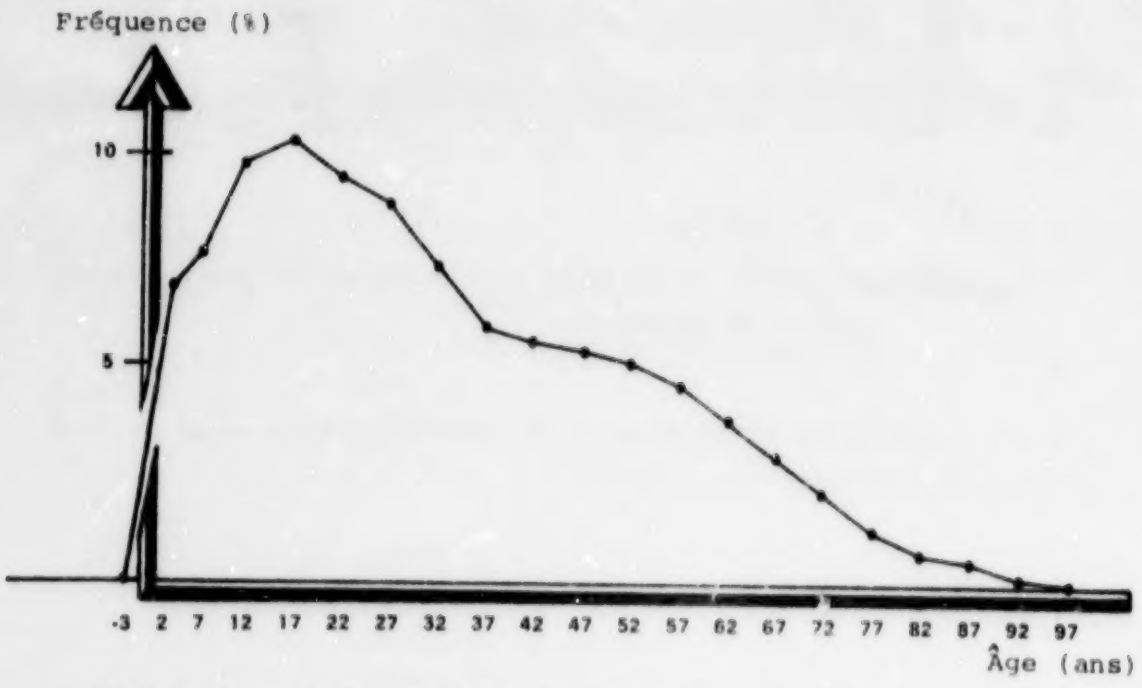


Figure 4 La population du Québec selon l'âge, 1976

SOURCE: Recensement du Canada - 1976

4. Polygone des effectifs cumulés ou des fréquences cumulées

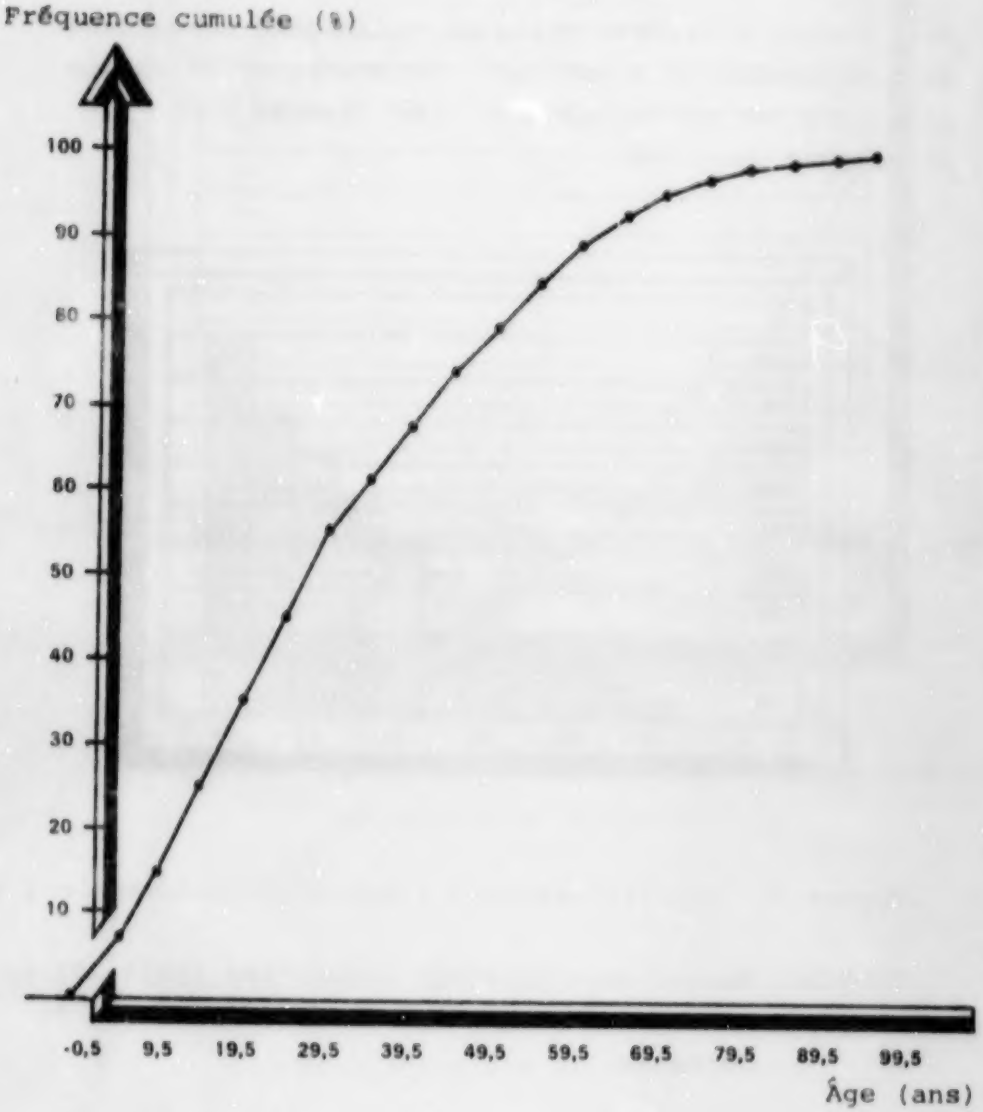


Figure 5 La population du Québec selon l'âge, 1976

5. Représentations diverses

Nous venons d'exposer quelques techniques classiques de représentation graphique. En pratique, il existe plusieurs autres techniques. Les figures 5 et 6 en donnent une idée.

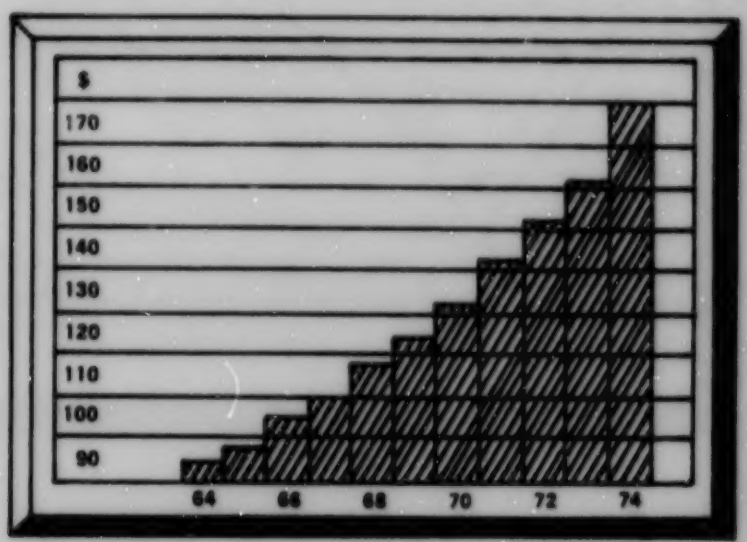


Figure 6 Les traitements et salaires hebdomadaires moyens

SOURCE: Survol de l'économie québécoise (1974-75) publié par le ministère de l'Industrie et du Commerce

Ce graphique est plutôt tendancieux. Au premier coup d'oeil, les traitements et salaires de 1974 semblent être 10 fois plus élevés qu'en 64 alors qu'en réalité ils ont à peine doublé.

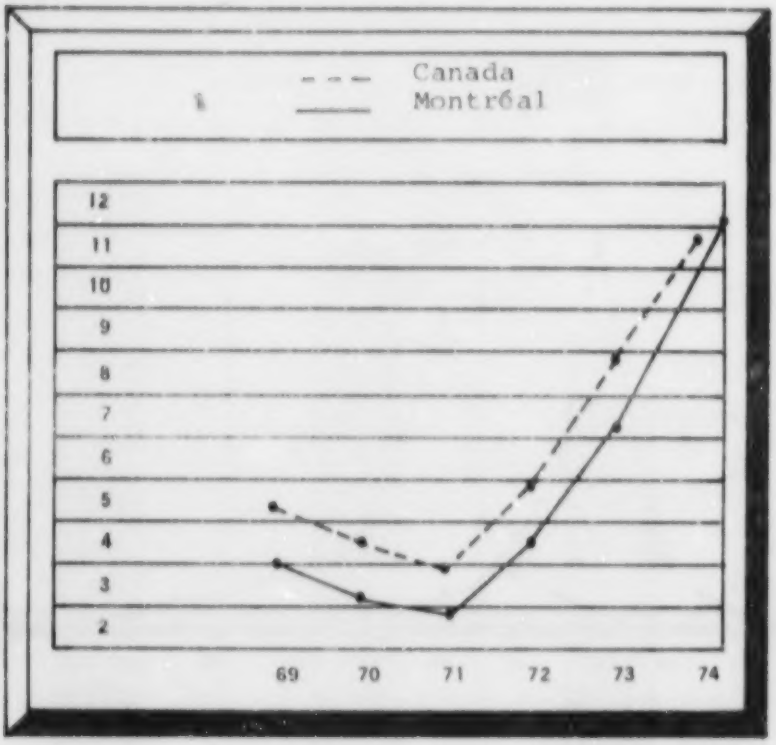


Figure 7 L'indice global des prix à la consommation
variations annuelles

SOURCE: Survol de l'économie québécoise
(1974-75) publié par le ministère
de l'Industrie et du Commerce

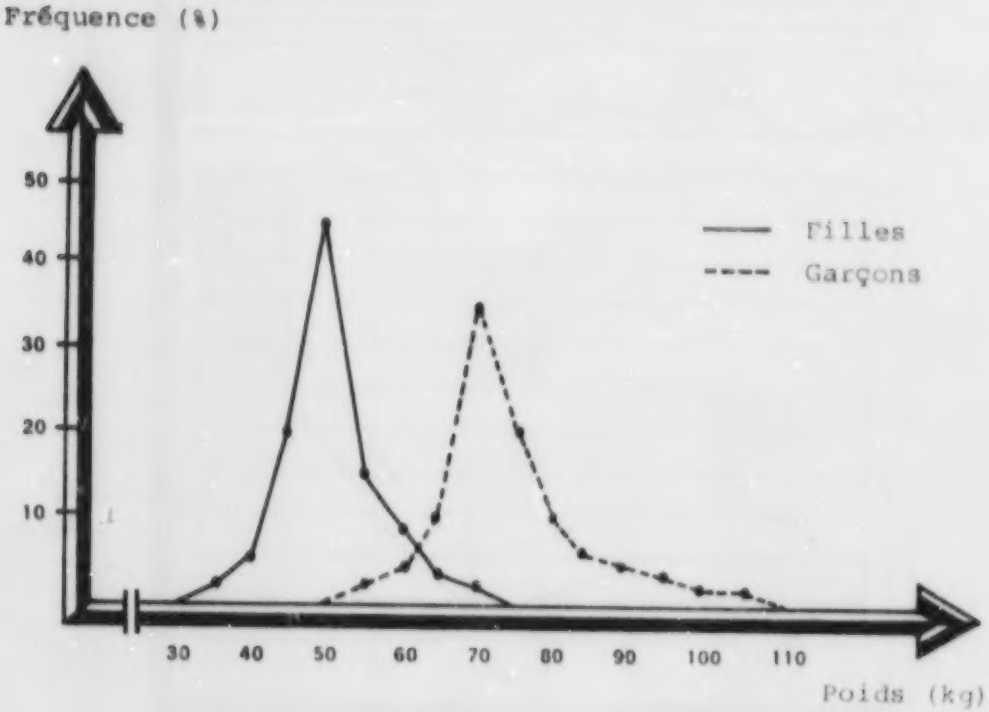


Figure 8 Répartition du poids d'un groupe de personnes selon le sexe.

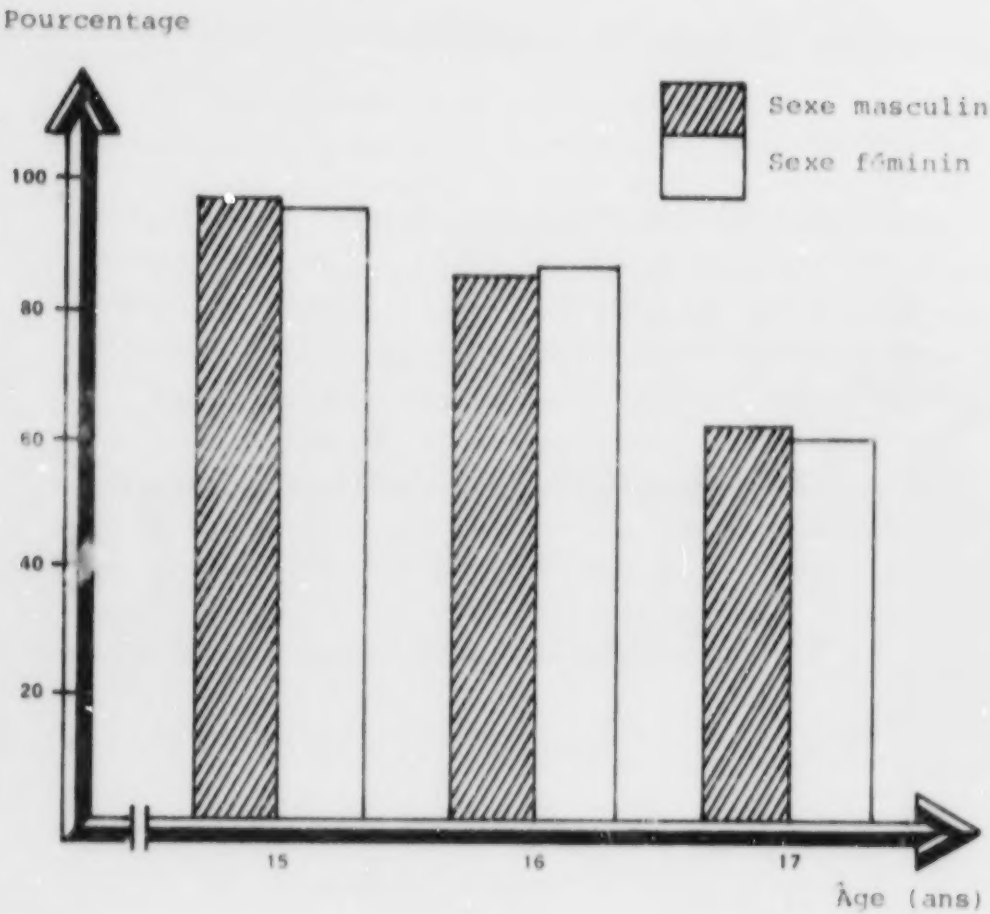


Figure 9 Pourcentage de la population canadienne âgée de 15 ans à 17 ans fréquentant l'école, selon le sexe

Source: Perspectives Canada II, recueil de statistiques sociales 1977

Comment mesurer autrement la dispersion?

1) Variance

C'est un troisième outil pour mesurer la dispersion. Il est analogue à l'écart moyen. Pour des raisons analytiques, au lieu de faire la moyenne des écarts comme dans l'écart moyen, *on calcule la moyenne des carrés des écarts.*

$$\text{Var.} = \frac{\text{Somme des carrés des écarts}}{\text{Nombre de données}}$$

2) Écart-type

C'est la racine carrée de la variance.

$$\text{Écart-type} = \sqrt{\text{variance}}$$

Comment représenter une distribution par un modèle théorique?

La nature offre parfois des exemples de régularités qui peuvent paraître étonnantes. Par exemple les tailles des adultes de sexe masculin semblent se répartir selon une courbe en forme de cloche. Ces régularités apparaissent dans les phénomènes les plus divers. Elles conduisent à penser que la distribution d'un groupe de données n'est pas quelconque mais qu'elle obéit de façon plus ou moins étroite à un schème que nous appelons loi statistique ou modèle théorique. Les modèles théoriques peuvent être exprimés par une formule mathématique, et ne dépendent que d'un petit nombre de paramètres.

A. Modèle pour les variables discrètes

Loi binômiale B(n;p)

1. Description de la structure

i) Expression analytique

$$f(x = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad 0 \leq k \leq n$$

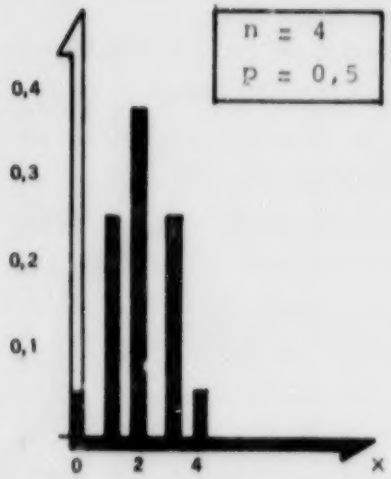
n et k des entiers non négatifs

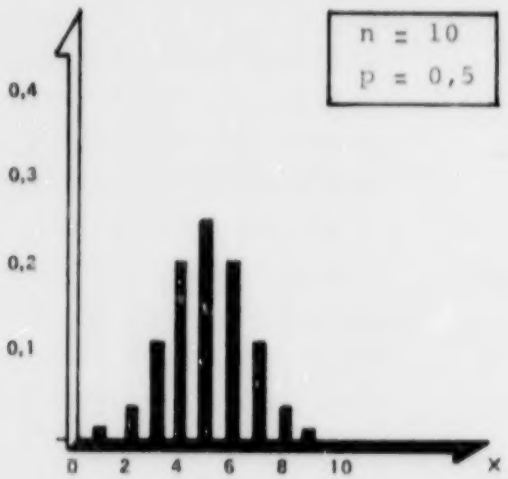
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

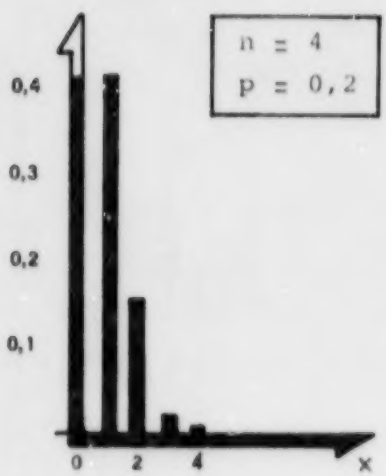
$$n! = n(n-1)(n-2)\ldots(2)(1)$$
$$0! = 1$$

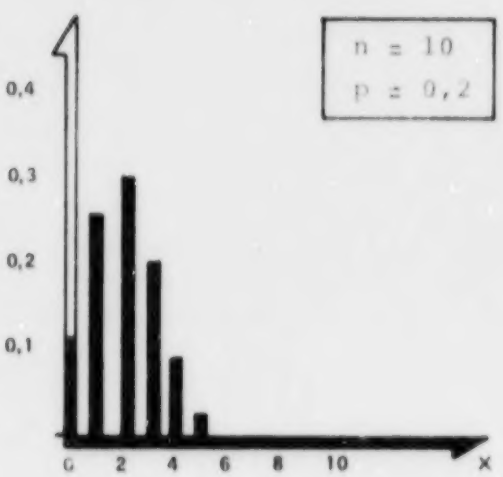
ii) Graphique

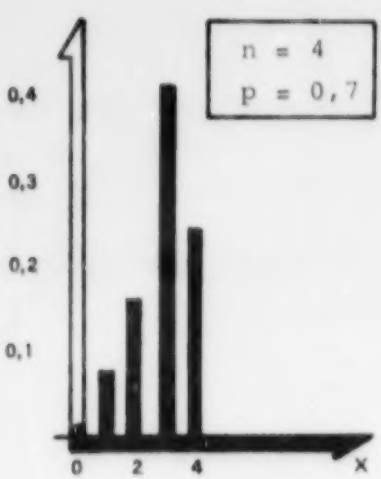
La forme du graphique n'est pas unique, elle varie en fonction de n et p.

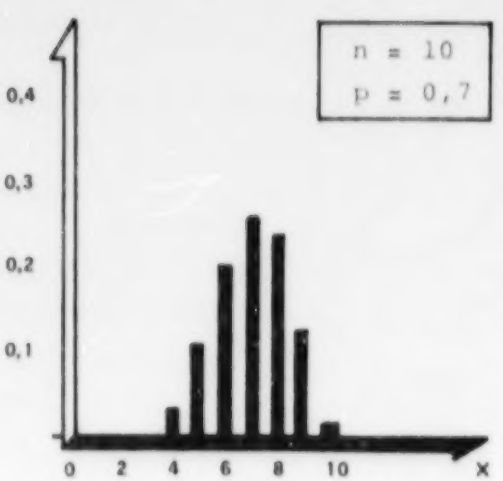

n = 4
p = 0,5
0,4
0,3
0,2
0,1
0
2
4
x


n = 10
p = 0,5
0,4
0,3
0,2
0,1
0
2
4
6
8
10
x


n = 4
p = 0,2
0,4
0,3
0,2
0,1
0
2
4
x


n = 10
p = 0,2
0,4
0,3
0,2
0,1
0
2
4
6
8
10
x


n = 4
p = 0,7
0,4
0,3
0,2
0,1
0
2
4
x


n = 10
p = 0,7
0,4
0,3
0,2
0,1
0
2
4
6
8
10
x

iii) Caractéristiques

Les caractéristiques varient en fonction de p.

	$p < \frac{1}{2}$	$p = \frac{1}{2}$	$p > \frac{1}{2}$
Pic			
- nombre	1	1	1
- localisation	à gauche	au centre	à droite
Symétrie par rapport au pic	non	oui	non

- il existe une famille de lois binômiales qui ne sont distinctes l'une de l'autre que par les valeurs de p et/ou n.

2. Synthèse de la structure

Moyenne de X = np

Variance de X = np(1-p)

3. Ajustement d'une distribution empirique.

Pour savoir si une variable observée suit approximativement une loi binômiale, on doit:

i) vérifier si les valeurs distinctes de la variable sont 0, 1, 2, ..., n,

ii) déterminer la valeur de n,

iii) calculer la moyenne,

iv) déterminer la valeur de p,

v) vérifier si la fréquence de chaque valeur différente de la variable vaut approximativement:

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

L'ajustement est d'autant meilleur que les différences entre les fréquences observées et les fréquences théoriques sont petites.

B. Modèles pour les variables continues

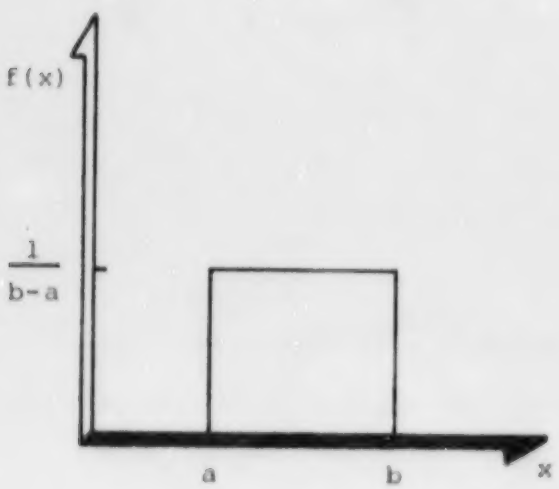
a) Loi uniforme

1) Description de la structure

i) Expression analytique

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

ii) Représentation graphique



iii) Caractéristiques

- aucun pic
- symétrie par rapport à la tendance centrale

2. Synthèse de la structure.

$$\text{Moyenne de } X = \frac{(a + b)}{2}$$

$$\text{Variance de } X = \frac{(b - a)^2}{12}$$

3. Ajustement d'une distribution empirique.

Pour savoir si une variable observée suit approximativement une loi uniforme, on doit:

i) vérifier si les valeurs distinctes de la variable sont consécutives,

ii) vérifier si la fréquence de chaque valeur différente de la variable est approximativement identique aux autres fréquences.

b) Loi normale $\mathcal{N}$ (moyenne; écart-type)

1) Description de la structure

i) Expression analytique

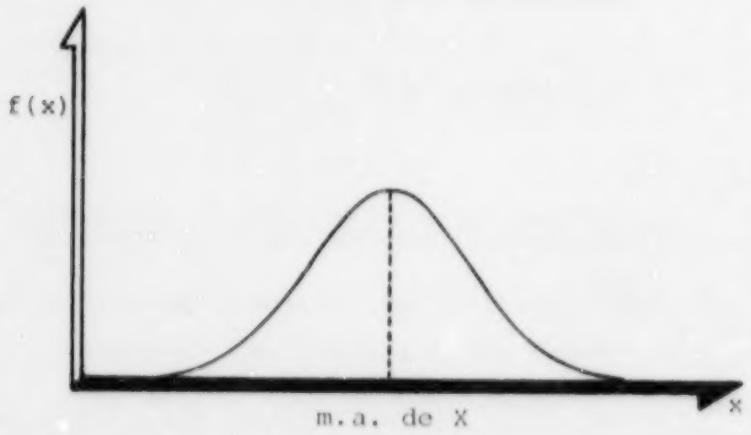
$$f(x) = \frac{1}{\sqrt{2\pi}\ (\text{é.-t. de } X)}\ e^{-\frac{1}{2}\left(\frac{x - \text{m.a. de } X}{\text{é.-t. de } X}\right)^2}$$

$$-\infty < x < \infty$$

où $\pi = 3,14159....$

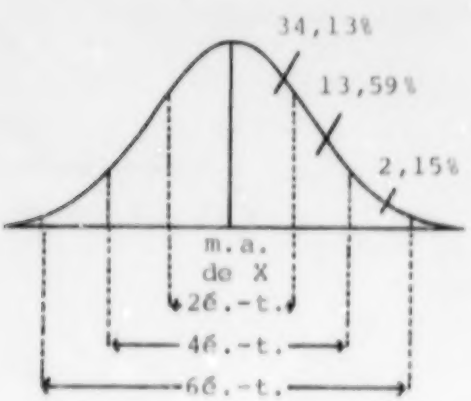
$e = 2,708....$

ii) Représentation graphique



iii) Caractéristiques

- un pic central
- symétrie par rapport à ce pic central
- courbe en forme de cloche
- pourcentage



- il existe une famille de lois normales distinctes l'une de l'autre seulement par leur moyenne et/ou leur variance.

2. Synthèse de la structure

Moyenne de X = μ ou $\bar{x}$

Variance de X = σ^2 ou s^2

3. Ajustement d'une distribution empirique

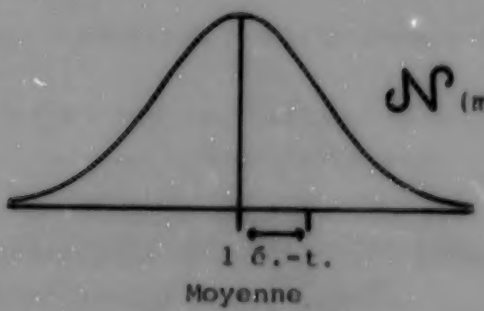
Pour vérifier si une variable observée suit approximativement une loi normale, on doit:

i) calculer sa moyenne et son écart-type,

ii) vérifier si les pourcentages donnés préalablement sont approximativement respectés.

Comment travailler avec une distribution normale?

Rappelons certaines caractéristiques importantes de la distribution normale.



- La courbe s'étend indéfiniment dans les deux directions tout en s'approchant de l'axe horizontal.

- La plus grande partie de la surface sous la courbe se trouve entre -4 é.-t. et + 4 é.-t. de la moyenne.

  L'aire sous la courbe vaut 1.

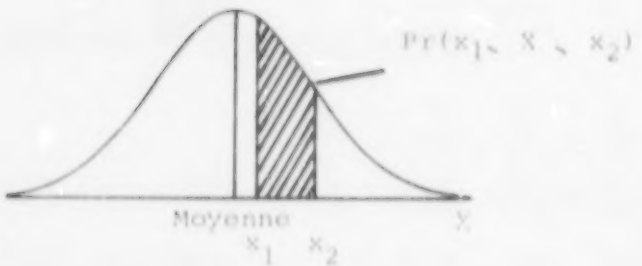
- Le pic de la courbe et la tendance centrale (moyenne, médiane, mode) se trouvent au centre.

- La courbe est symétrique par rapport à l'axe central vertical.

- La probabilité qu'un résultat se situe entre deux valeurs x_1 et x_2 est égale à:

$$\frac{\text{aire sous la courbe dans cet intervalle}}{\text{aire totale}}$$

ce qui donne toujours un résultat qui se situe entre 0 et 1.

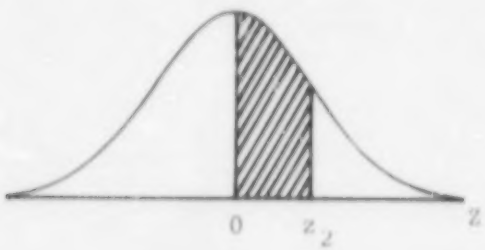


En pratique, la valeur de l'aire ($Pr(x_1 \leq X \leq x_2)$) à partir de l'équation exige un calcul long et complexe. Heureusement, on peut obtenir cette valeur à partir d'une table.

1. La distribution normale centrée, réduite $\mathcal{N}(0;1)$

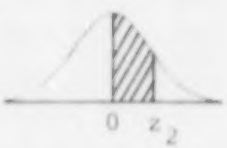
Utilisation de la table

La table donne l'aire de la région sous la courbe $\mathcal{N}(0;1)$ comprise entre les points 0 et z_2 où $z_2 \in [0;3,99]$.



On représente cette aire par $A_0^{z_2}$

L'aire, $A_0^{z_2}$, sous la courbe $\mathcal{N}(0;1)$ est égale à la probabilité que la variable Z soit comprise dans l'intervalle $[0;z_2]$

Donc $A_0^{z_2} = Pr(0 \leq Z \leq z_2) =$ 

On constate que $A_{-\infty}^{0} = A_0^{\infty} = 0,5$

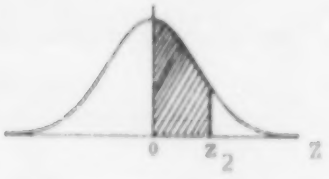
En effet

1) $A_{-\infty}^{\infty} =$ $= 1$

2) $A_{-\infty}^{0} =$ $= A_0^{\infty} =$ $= 0,5$

(par symétrie)

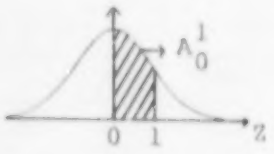
AIRES
limitées par la
COURBE NORMALE
CENTREE REDUITE
de 0 à z_2



z_2	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

a) Recherche d'une aire sous la courbe

i) Directement dans la table de la $\mathcal{N}(0;1)$.



$A_0^1 = Pr(0 < Z < 1) = 0,3413$

= aire sous la courbe $\mathcal{N}(0;1)$ comprise entre les points 0 et 1. Donc cette aire est directement donnée dans la table.

ii) Indirectement dans la table de la $\mathcal{N}(0;1)$.

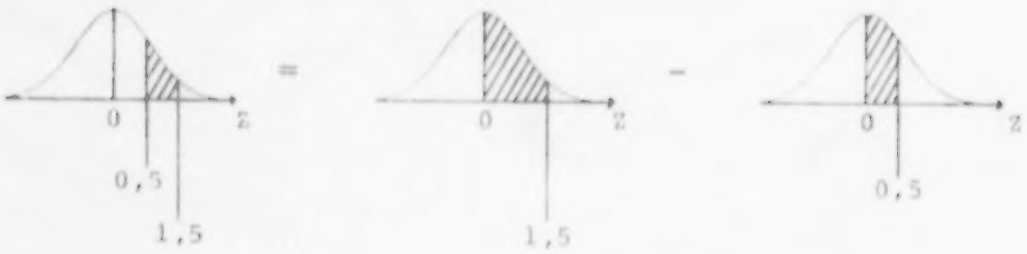
1) *Par symétrie*

$Pr(-1 < Z < 0) = A_{-1}^0 = A_0^1 = 0,3413$

Pour toutes valeurs de z_2 on a

$$A_{-z_2}^0 = A_0^{z_2}$$

2) *Par soustraction*

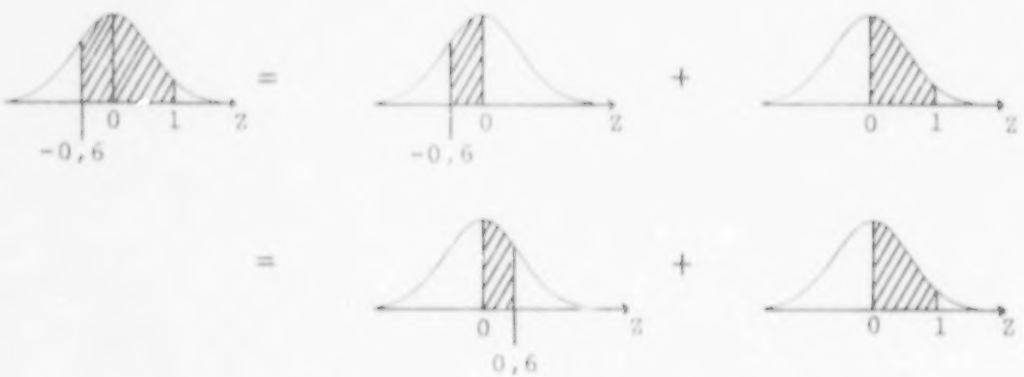


$$Pr(0,5 \leq Z \leq 1,5) = A_{0,5}^{1,5} = A_0^{1,5} - A_0^{0,5}$$

$$= 0,4332 - 0,1915$$

$$= 0,2417$$

3) *Par symétrie et addition*



$$Pr(-0,6 \leq Z \leq 1) = A_{-0,6}^{1} = A_0^{0,6} + A_0^{1}$$

$$= 0,2258 + 0,3413$$

$$= 0,5671$$

4) *Par symétrie et soustraction*

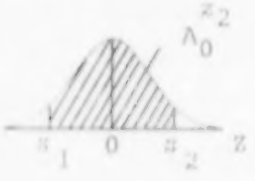
$$Pr(-\infty < Z < -1) = A_{-\infty}^{-1} = A_0^{\infty} - A_0^{1}$$

$$= 0,5 - 0,3413$$

$$= 0,1587$$

b) On peut utiliser cette table pour évaluer une valeur de z_2 à partir d'une aire connue.

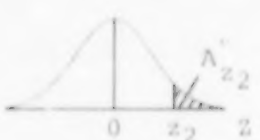
1) Quelles sont les valeurs d'abcisse qui donnent une aire de 95% au centre de la distribution?



$A_{z_1}^{z_2} = 0,95 \Rightarrow A_{z_1}^{0} = A_0^{z_2} = 0,475$ $\quad z_2 = 1,96$ (dans la table)

$z_1 = -z_2 = -1,96$

2) Quelle est la valeur d'abcisse qui donne une aire de 5% complètement à droite de la distribution?

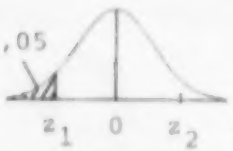


$A_{z_2}^{\infty} = 0,05 \Rightarrow A_0^{z_2} = 0,45$

$\Rightarrow z_2 = 1,645$

3) Quelle est la valeur d'abcisse qui donne une aire de 5% complètement à gauche de la distribution?

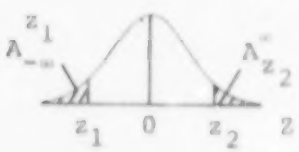
$A_{-\infty}^{z_1} = 0,05$



$$A_{-\infty}^{z_1} = 0,05 \Rightarrow A_{z_1}^{0} = 0,45 \Rightarrow A_{0}^{z_2} = 0,45$$

$$\Rightarrow z_1 = -z_2 = -1,645$$

4) Quelles sont les valeurs d'abcisse qui donnent une aire de 5% répartie également à chaque extrémité de la distribution?



$$A_{-\infty}^{z_1} + A_{z_2}^{\infty} = 0,05 \Rightarrow A_{-\infty}^{z_1} = A_{z_2}^{\infty} = 0,025$$

$$A_{z_1}^{0} = A_{0}^{z_2} = 0,475$$

$$z_2 = 1,96 \quad \text{et} \quad z_1 = -1,96$$

2. La distribution normale quelconque $\mathcal{N}$(moyenne; écart-type)

Que fait-on, si on a affaire à une distribution normale de moyenne différente de 0 et d'écart-type différent de 1?

Il faut transformer cette distribution normale en une $\mathcal{N}(0;1)$, à l'aide de la variable Z.

$$Z = \frac{X - \text{moyenne de } X}{\text{écart-type de } X}$$

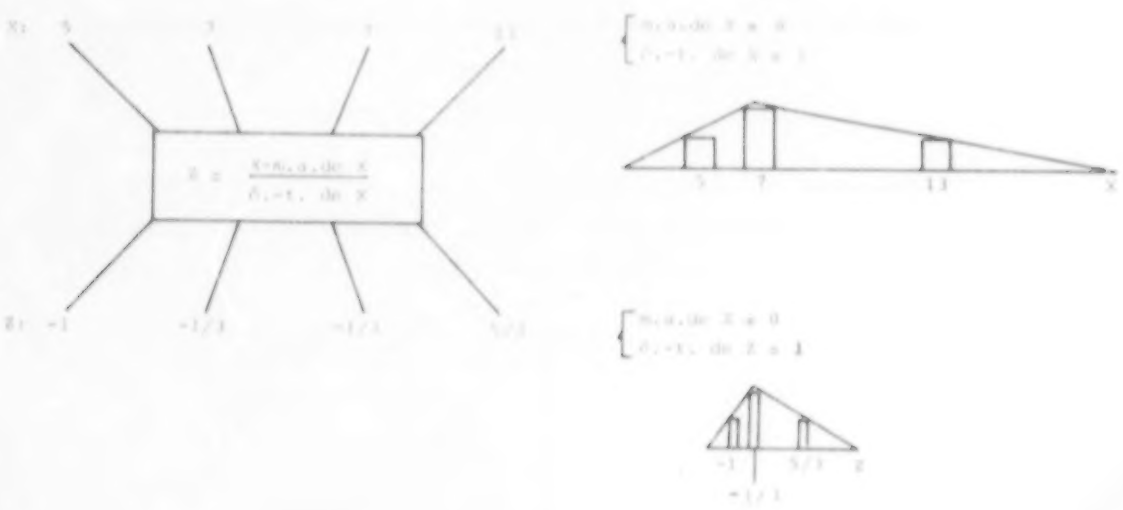
Cette transformation a pour effet:

1) de supprimer la valeur (Z est un nombre pur);

2) de déplacer la moyenne à 0;

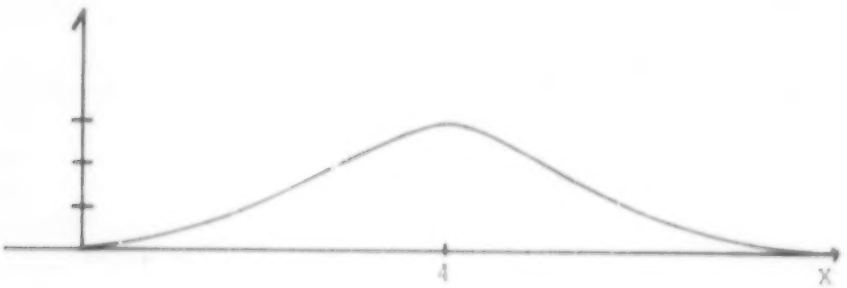
3) d'ajuster la dispersion des points pour obtenir un écart-type égal à 1.

Exemple explicatif de cette formule:

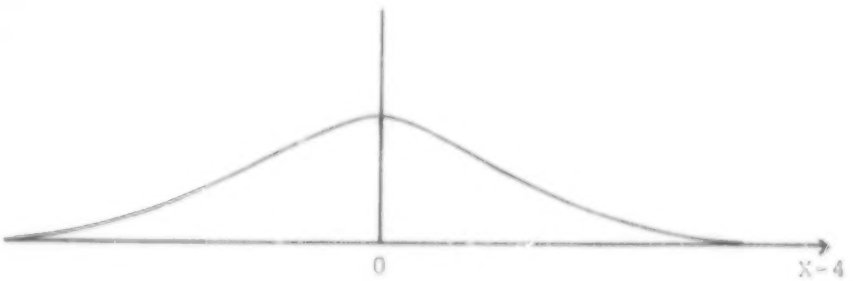


Représentation graphique de l'effet de la cote Z sur la courbe normale.

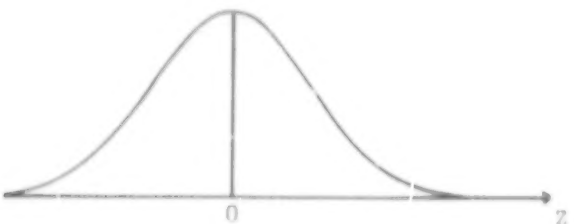
Soit une variable normale X dont la moyenne est 4 et l'écart-type vaut 2.



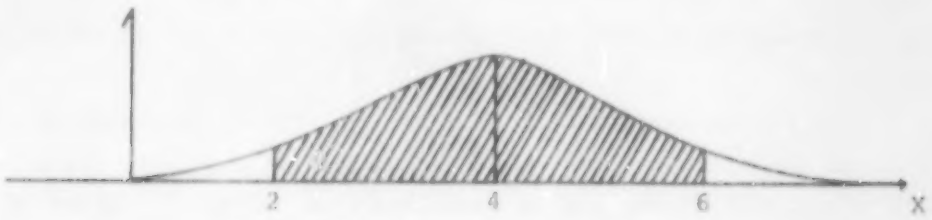
La soustraction de la moyenne 4 à chaque valeur de X produit un déplacement de la courbe de 4 unités vers la gauche. La dispersion ne change pas.



La division de la variable X-4 par son écart-type 2 produit un rétrécissement de la courbe. La dispersion diminue.



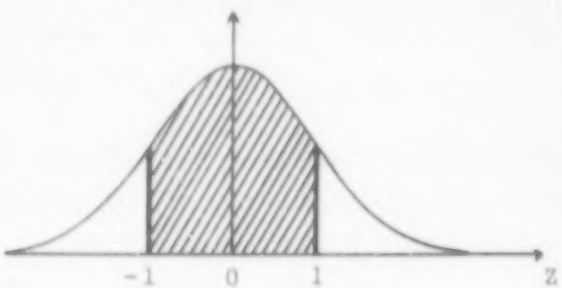
La cote Z préserve les surfaces sous la courbe normale. La $Pr(2 < X < 4)$ est représentée par la surface hachurée suivante.



Il est possible de montrer que

$$Pr(2 < X < 4) = Pr(\frac{2-4}{2} < \frac{X-4}{2} < \frac{6-4}{2})$$
$$= Pr(-1 < Z < 1)$$

La dernière probabilité est représentée par la surface hachurée suivante.



Les deux surfaces hachurées sont donc égales.

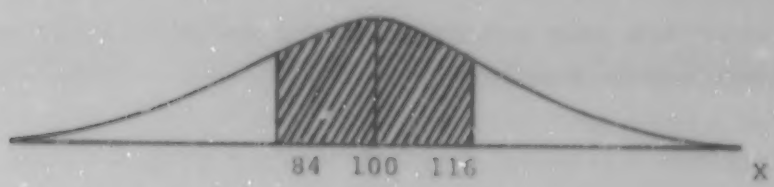
Donc toutes les valeurs en X d'une distribution normale peuvent être converties en valeurs de Z et on revient alors à une $\mathcal{N}(0;1)$ et pour les calculs on emploie en conséquence la même table.

Voici un exemple, qui montre comment ceci peut servir:

Soit X le quotient intellectuel d'une personne choisie au hasard dans une population où le quotient intellectuel est distribué normalement de moyenne 100 et d'écart-type 16. Supposons que l'on veuille calculer la probabilité

$$Pr(84 < X < 116)$$

C'est-à-dire la probabilité que la personne choisie ait un quotient intellectuel entre 84 et 116.



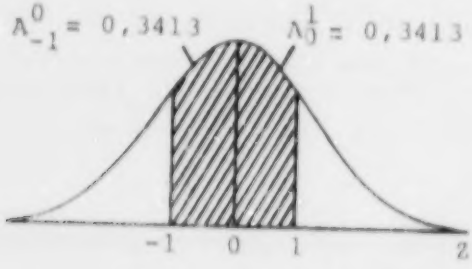
Nous exprimons cet événement en termes de la variable Z, de la façon suivante:

$$\frac{84-100}{16} < \frac{X-100}{16} < \frac{116-100}{16}$$

pour obtenir $-1 < Z < 1$

alors $Pr(84 < X < 116) = Pr(-1 < Z < 1)$

Nous pouvons calculer cette probabilité en trouvant l'aire sous la courbe $\mathcal{N}(0;1)$ entre -1 et 1



$$A_{-1}^{1} = 2(0,3413) = 0,6826$$

Comment trouver la structure conjointe de deux caractéristiques?

Pour trouver la structure conjointe de deux caractéristiques, il faut:

1) Trouver la distribution conjointe des deux caractéristiques, c'est-à-dire en combinant chaque catégorie d'une caractéristique à chaque catégorie de l'autre caractéristique, trouver l'effectif (ou la fréquence) de ces combinaisons.

2) Représenter la distribution conjointe des deux caractéristiques par un des moyens suivants:

   i) Le tableau 6 représente la distribution conjointe des caractéristiques taille (en cm) et longueur de ceinture (en cm) pour un groupe de 20 personnes.

      Dans une distribution conjointe, la distribution d'une seule caractéristique à la fois s'appelle la distribution marginale de cette caractéristique. Ainsi, dans l'exemple précédent, la distribution mar-

ginale de la taille peut être représentée par le tableau 7.

Tableau 6 La répartition de 20 personnes selon la taille et la longueur de ceinture

Taille (cm) / Longueur de ceinture (cm)	150	160	170	180	Total
70	2	1	1		4
75		5	1	1	7
80		2	3	1	6
85			1	2	3
Total	2	8	6	4	20

Tableau 7 La répartition de 20 personnes selon la taille

Taille (cm)	Effectif
150	2
160	8
170	6
180	4
Total	20

ii) Nuage de points

Pour construire un nuage de points, il s'agit de situer les catégories d'une caractéristique sur l'abcisse et les catégories de l'autre caractéristique sur l'ordonnée puis représenter chaque paire de données par un point.

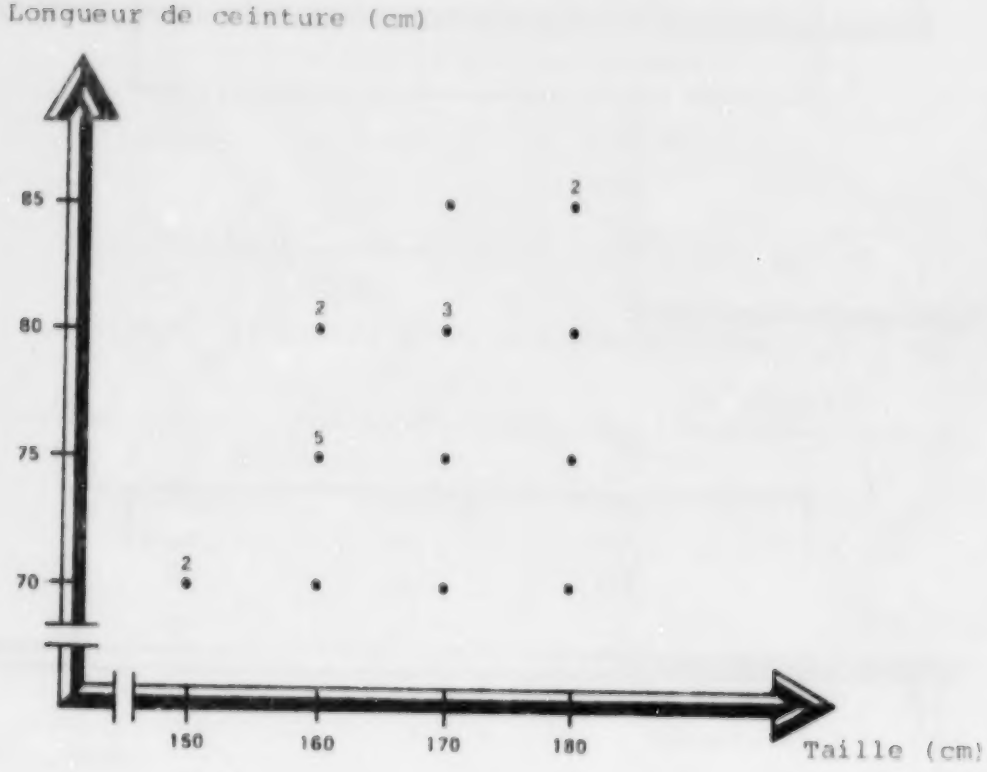


Figure 10 La répartition de 20 personnes selon la taille et la longueur de ceinture

Les chiffres près des points indiquent le nombre de paires de données qui se trouvent à cet endroit. On peut aussi faire un point proportionnel à l'effectif.

3) Chercher les liens entre les 2 caractéristiques.

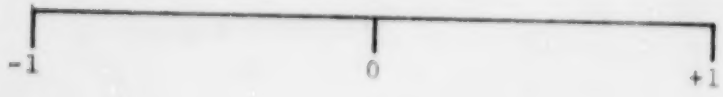
Les questions à répondre:

- *existe-t-il un lien?*
- *le genre de lien?*
- *le degré du lien?*
- *l'expression algébrique du lien?*

1- Le coefficient de corrélation linéaire (r)

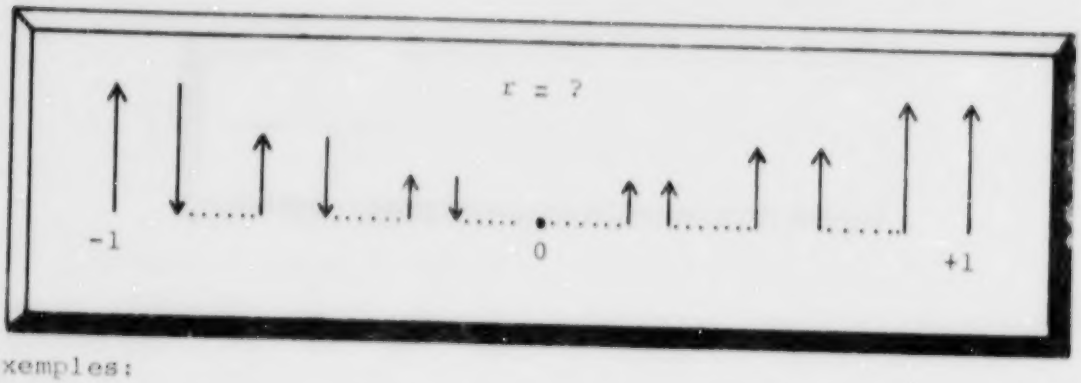
L'outil répond aux trois premières questions.

Cet outil est gradué de -1 à +1.



- Existence du lien	$r \neq 0$	
- Genre de lien	$r < 0$	(Les valeurs des caractéristiques s'opposent.)
	$r > 0$	(Les valeurs des caractéristiques concordent.)
- Degré du lien	$r \longrightarrow +1$ ou $r \longrightarrow -1$	(Le lien est fort.)
	$r \longrightarrow 0$	(Le lien est faible.)

D'une façon imagée

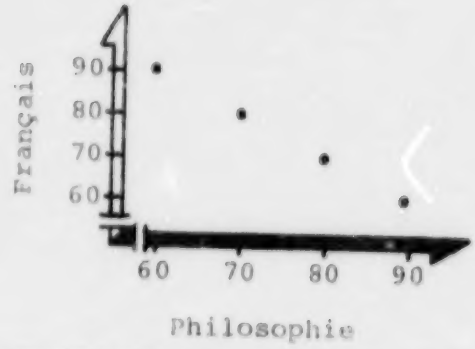


Exemples:

a)

	Notes en Français	Philosophie
Pierre	90	60
Paul	80	70
Jean	70	80
Thomas	60	90

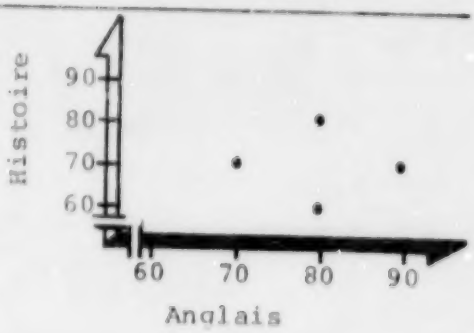
(r = -1)



b)

	Notes en Histoire	Anglais
Pierre	60	80
Paul	70	70
Jean	70	90
Thomas	80	80

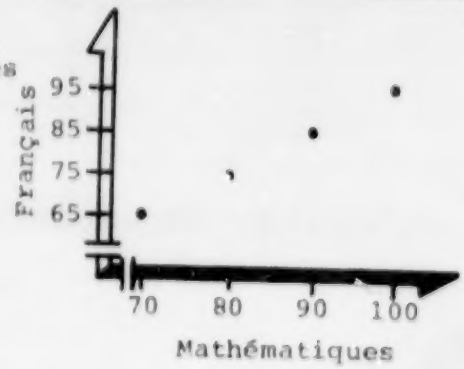
(r = 0)



c)

	Notes en Français	Mathématiques
Pierre	95	100
Paul	85	90
Jean	75	80
Thomas	65	70

(r = +1)



Expression algébrique de l'outil

$$r = \frac{\text{Cov}(x_i, y_i)}{(\text{é.-t. des } x_i)(\text{é.-t. des } y_i)}$$

x_i : donnée de la première caractéristique

$\text{cov}(x_i, y_i)$: covariance des données des deux caractéristiques

é.-t. des x_i : écart-type des données de la première caractéristique

y_i : donnée de la deuxième caractéristique

é.-t. des y_i : écart-type des données de la deuxième caractéristique

Exemple:

	Anglais	Géographie
Pierre	70	80
Paul	80	50
Jean	60	90
Thomas	90	70

x_i - m.a. des x_i	y_i - m.a. des y_i	$(x_i$ - m.a. des $x_i)$ $(y_i$ - m.a. des $y_i)$
70-75	80-72,5	-37,5
80-75	50-72,5	-112,5
60-75	90-72,5	-262,5
90-75	70-72,5	-37,5
		-450

$$\text{Cov}(x_i, y_i) = \frac{-450}{4}$$

$$r = \frac{-450}{(11,18) \times (14,79) \times 4} = -0,68$$

2- La droite de régression

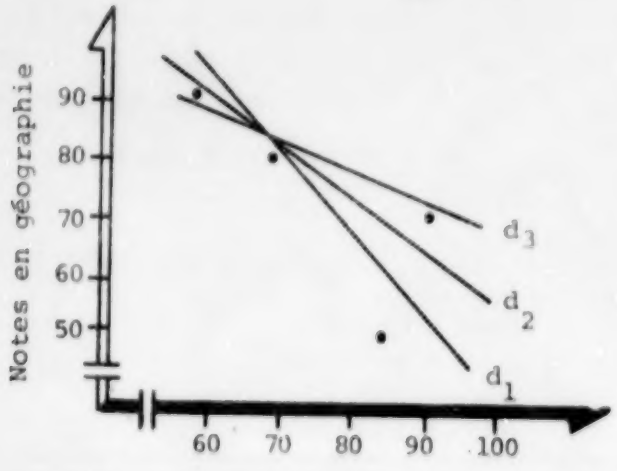
Dans l'exemple précédent, on trouve que:

- il existe un lien entre les notes en anglais et les notes en géographie ($r \neq 0$);
- les valeurs des caractéristiques s'opposent c'est-à-dir l'étudiant qui a des notes fortes dans une matière tend à avoir des notes faibles dans l'autre matière et vice versa ($r \simeq -0,68 < 0$);
- le lien n'est pas très fort.

Il nous reste à exprimer ce lien. On suppose qu'il est linéaire c'est-à-dire qu'il peut s'exprimer par une relation de la forme.

$$Y = mX + b$$

Si on fait le graphique des notes, on obtient des points non alignés.



On peut ainsi passer plusieurs droites (d_1, d_2, d_3,...) par ces points. Laquelle exprime le mieux la relation entre nos deux caractéristiques.

Le principe des moindres carrés nous permet d'obtenir l'équation qui explique le mieux Y en fonction de X.

$$Y = mX + b$$

où

$$m = \frac{\text{Cov}(x_i, y_i)}{\text{variance des } x_i}$$

$$b = \text{m.a.des } y_i - m(\text{m.a.des } x_i)$$

	Anglais (x_i)	Géographie (y_i)
Pierre	70	80
Paul	80	50
Jean	60	90
Thomas	90	70
	$\sum x_i = 300$	$\sum y_i = 290$

$$m = \frac{-450}{125 \times 4} = -0,9$$

$$b = \frac{290}{4} - (-0,9) \times \frac{300}{4} = 140$$

$$Y = -0,9\,X + 140$$

Notes en géographie = -0,9 x Notes en anglais + 140

Comment choisir un échantillon d'une population?

Il y a plusieurs méthodes d'échantillonnage:

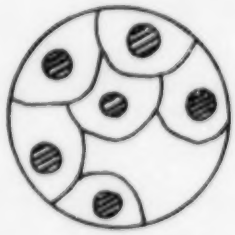
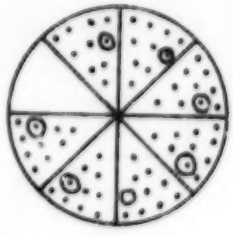
- *échantillonnage aléatoire simple,*
- *échantillonnage systématique,*
- *échantillonnage stratifié aléatoire,*
- *échantillonnage par grappes,*
  (ou à un degré)
- *échantillonnage à 2, 3, ..., n degrés.*

Le choix de la méthode dépendra toujours du problème à traiter.

Cependant, quelle que soit la méthode, le choix des éléments de l'échantillon se fait d'une façon aléatoire et un élément de la population peut ou ne pas réapparaître dans l'échantillon selon que le tirage est fait avec ou sans remise.

La différence entre les méthodes provient des diverses façons de diviser ou de subdiviser une population en fonction des besoins de l'étude.

Tableau comparatif des méthodes d'échantillonnage

Méthode	Division ou subdivision de la population		Construction de l'échantillon
Échantillonnage aléatoire simple	- aucune		Éléments choisis dans toute la population
Échantillonnage stratifié aléatoire	- partitionnée en sous-groupes homogènes		Éléments choisis au hasard dans chaque sous-groupe de la population
Échantillonnage par grappes (ou à 1 degré)	- divisée en sous-groupes représentatifs de la population		Éléments des sous-groupes choisis au hasard
Échantillonnage à 2,3,..., n degrés	- divisée en sous-groupes représentatifs de la population		Éléments choisis au hasard dans chaque sous-groupe choisi au nasard
Échantillonnage systématique	- partitionnée en sous-groupes de même taille		Choix aléatoire du premier élément seulement

Comment choisir les éléments d'un échantillon d'une façon aléatoire?

On doit:

1) Établir une base de sondage;

(c'est-à-dire faire la liste des éléments qui serviront au choix aléatoire des éléments de l'échantillon.)

2) Se servir de techniques connues comme les fonctions pseudo-aléatoires, les tables de nombres aléatoires.

a) Les fonctions pseudo-aléatoires

Les fonctions qui ont été programmées sur ordinateur pour simuler le hasard. Pour les employer, il faut suivre les instructions de chaque constructeur.

b) Les tables de nombres aléatoires

C'est une série de chiffres qui obéissent à la loi discrète et uniforme.

$$L(x) = \frac{1}{10} \qquad 0 \leqslant x \leqslant 9$$

FISHER, R.A. Statistical tables for biological, agricultural and medical research. Hafner et co., New York, 1970.

Nombres au hasard

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 34 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 94 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 49 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 91	00 52 43 48 85	27 55 26 89 62
56 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	15 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 75 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	42 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43
17 53 77 58 71	71 41 61 50 72	12 41 94 96 26	44 95 27 36 99	02 96 74 30 83
90 26 59 21 19	23 52 23 33 12	96 93 02 18 39	07 02 18 36 07	25 99 32 70 23
41 23 52 55 99	31 04 49 69 96	10 47 48 45 88	13 41 43 89 20	97 17 14 49 17
60 20 50 81 69	31 99 73 68 68	35 81 33 03 76	24 30 12 48 60	18 99 10 72 34
91 25 38 05 90	94 58 28 41 36	45 37 59 03 09	90 35 57 29 12	82 62 54 65 60
34 50 57 74 37	98 80 33 00 91	09 77 93 19 82	74 94 80 04 04	45 07 31 66 49
85 22 04 39 43	73 81 53 94 79	33 62 46 86 28	08 31 54 46 31	53 94 13 38 47
09 79 13 77 48	73 82 97 22 21	05 03 27 24 83	72 89 44 05 60	35 80 39 94 88
88 75 80 18 14	22 95 75 42 49	39 32 82 22 49	02 48 07 70 37	16 04 61 67 87
90 96 23 70 00	39 00 03 06 90	55 85 78 38 36	94 37 30 69 32	90 89 00 76 33

GREMY, F. Salmon, D. Bases statistiques pour la recherche médicale et biologique. Dunod, Paris, 1969.

Exemple

Choisir d'une façon aléatoire 30 personnes d'une population de 400 personnes.

1) Établir la base de sondage

	Nombre		Nombre
P_1	001	P_{11}	011
P_2	002	P_{12}	012
P_3	003	.	.
P_4	004	.	.
P_5	005	.	.
P_6	006	.	.
P_7	007	.	.
P_8	008	.	.
P_9	009	.	.
P_{10}	010	P_{400}	400

2) Procédé

Lire la table de nombres au hasard par colonne ou par rangée de trois chiffres et prendre les 30 premiers chiffres entre 1 et 400.

Par rangée:

034 743 738 636 964 736 614 698 637

162 332 etc. à éliminer (400)

Pour aller plus vite, on aurait pu attribuer deux nombres à chaque personne.

	Nombres	
P_1	001	401
P_2	002	402
P_3	003	403
.	.	.
.	.	.
.	.	.
P_{400}	400	800

Ainsi les nombres 743 738 636 736 614 698 637 n'auraient pas été éliminés.

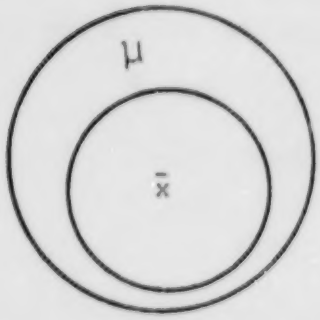
Combien d'éléments dans l'échantillon?

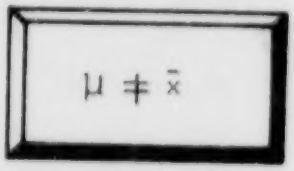
Combien faut-il d'éléments dans l'échantillon pour obtenir de l'information au sujet de la population selon la précision et le degré de confiance souhaités?

L'emploi d'un échantillon implique toujours une possibilité d'erreur; faute de pouvoir éliminer cette erreur, il faut l'amadouer, la maîtriser, la contrôler. On a fait nos premiers pas dans ce domaine à l'expérience 7.

Si on choisit tous les échantillons possibles d'une population:

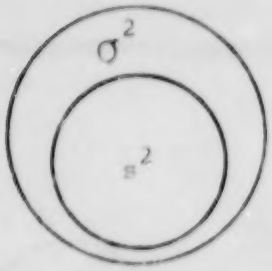
1) la moyenne d'un échantillon ($\bar{x}$) n'est pas très souvent égale à la moyenne de la population (μ):





(le plus souvent)

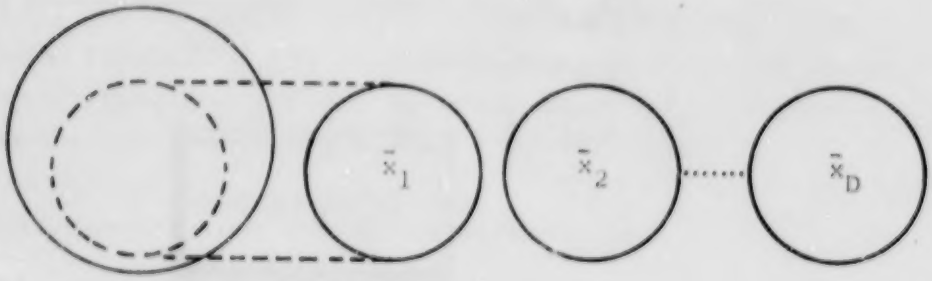
2) la variance d'un échantillon (s^2) n'est presque jamais égale à la variance de la population (σ^2);



$$\sigma^2 \neq s^2$$

(le plus souvent)

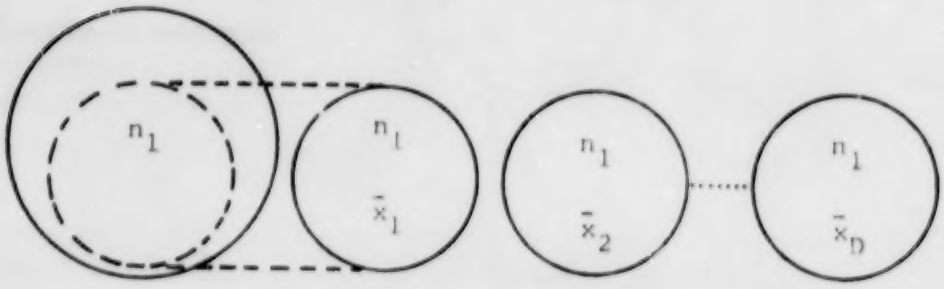
3) la moyenne des moyennes d'échantillons ($\mu_{\bar{x}}$) est égale à la moyenne de la population (μ);



$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_D}{D} = \mu$$

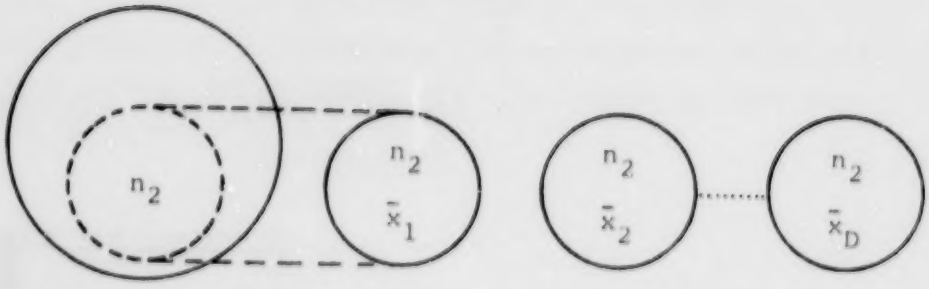
$$\mu_{\bar{x}} = \mu$$

4) la variance des moyennes d'échantillons ($\sigma^2_{\bar{x}}$) n'est pas égale à la variance de la population (σ^2) mais elle diminue avec l'augmentation de la taille de l'échantillon (n).



$$\sigma^2_{\bar{x}}(n_1) = \frac{(\bar{x}_1 - \mu_{\bar{x}})^2 + (\bar{x}_2 - \mu_{\bar{x}})^2 + \ldots + (\bar{x}_D - \mu_{\bar{x}})^2}{D}$$

$$\sigma^2_{\bar{x}}(n_1) \neq \sigma^2$$



$$\sigma^2_{\bar{x}}(n_2) = \frac{(\bar{x}_1 - \mu_{\bar{x}})^2 + (\bar{x}_2 - \mu_{\bar{x}})^2 + \ldots + (\bar{x}_D - \mu_{\bar{x}})^2}{D}$$

$$\sigma^2_{\bar{x}}(n_2) \neq \sigma^2$$

$$\sigma^2_{\bar{x}}\ (n_1) \cdot \sigma^2_{\bar{x}}\ (n_2) \qquad \text{si } n_1 \cdot n_2$$

Donc, la chance de retrouver la moyenne de l'échantillon ($\bar{x}$) autour de la moyenne de la population (μ) est de plus en plus élevée si on augmente la taille de l'échantillon (n).

Ainsi, on peut construire un échantillon dont l'erreur sur l'estimation de la moyenne de la population n'excède pas la valeur considérée comme acceptable dans le problème.

Cependant, quelle est la taille de cet échantillon? Pour y répondre facilement, il faut connaître en détail ce qu'on peut avoir lorsqu'on prélève un échantillon d'une population quelle que soit la taille de l'échantillon.

Heureusement, ces résultats sont connus.

En effet, lorsque la taille de l'échantillon est d'au moins 30 éléments, la distribution des moyennes d'échantillons suit une loi normale (voir le manuel d'instructions à ce sujet).

$$\bar{x} \sim \mathcal{N}(\mu_{\bar{x}} ; \sigma_{\bar{x}})$$

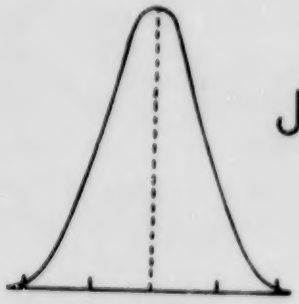
où $\mu_{\bar{x}} = \mu$

$$\sigma_{\bar{x}} = \begin{cases} \dfrac{\sigma}{\sqrt{n}} & \text{(tirage avec remise)} \\ \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}} & \text{(tirage sans remise)} \end{cases}$$

N: taille de la population

---

i) n_1



$\mathcal{N}(\mu_{\bar{x}} ; \sigma_{\bar{x}})$

$\mu_{\bar{x}} = \mu$

Distribution des moyennes d'échantillons

$n_1 > n_2$

ii) n_2

$\mathcal{N}(\mu_{\bar{x}} ; \sigma_{\bar{x}})$

$\mu_{\bar{x}} = \mu$

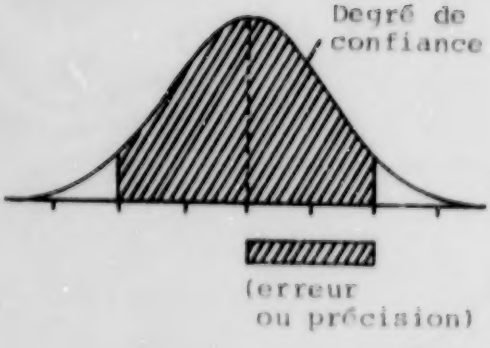
Distribution des moyennes d'échantillons

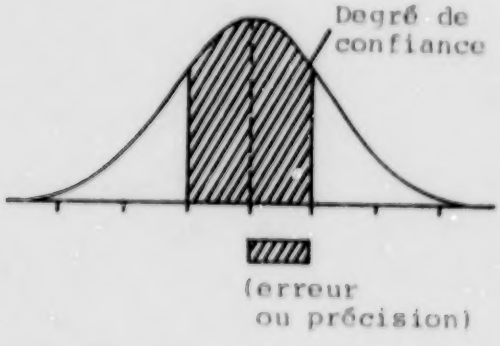
Étude de la précision et du degré de confiance en relation avec n.

1. Pour un n fixé,

- la précision (ou l'erreur acceptable) varie avec le degré de confiance et vice versa.

a) Plus on veut de précision (ou une petite erreur), moins on sera confiant de l'obtenir dans les résultats.



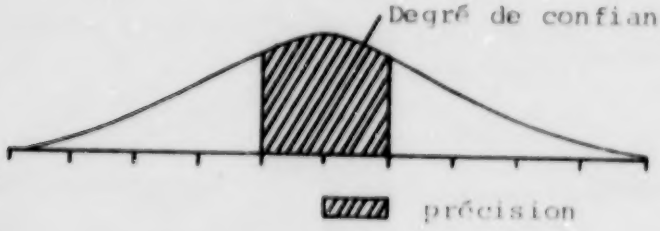


b) Plus on veut être confiant dans les résultats, moins on aura de précision (plus grande sera l'erreur d'estimation).

2. Pour une précision (ou erreur acceptable),

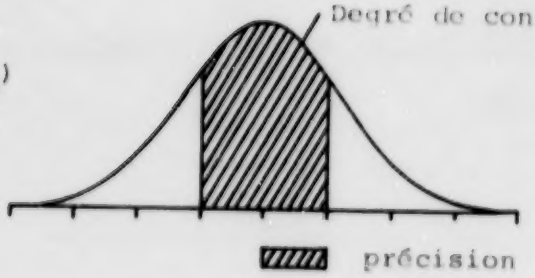
   - plus on augmente la taille de l'échantillon, plus la chance d'avoir cette précision dans l'estimation augmente (ou de ne pas commettre une erreur plus grande).

n_3



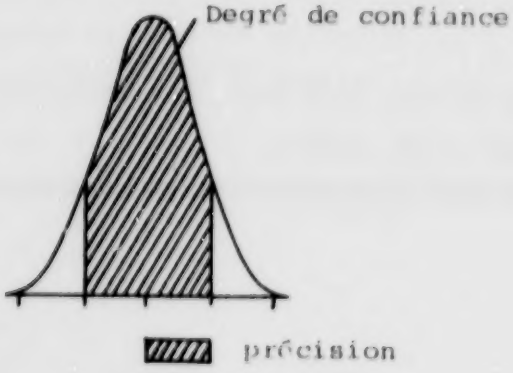
n_2

$(n_2 > n_3)$



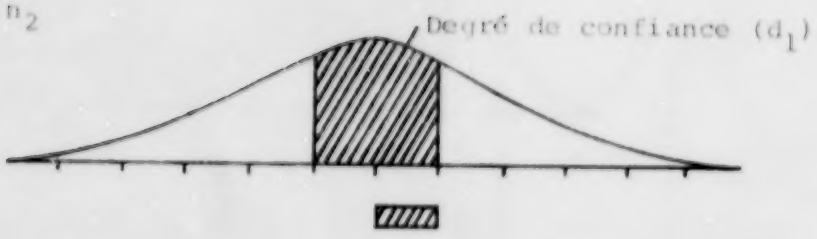
n_1

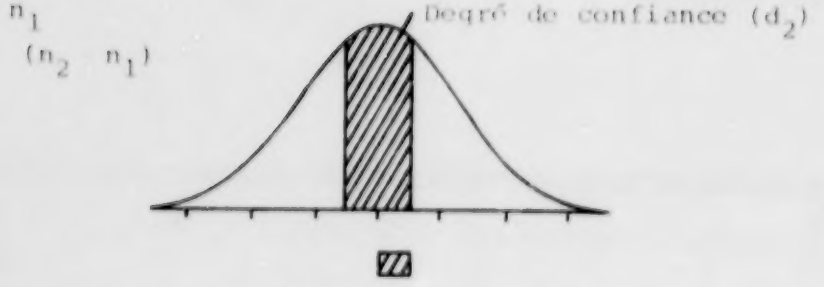
$(n_1 > n_2 > n_3)$



3. Pour un degré de confiance convenable,

   - plus on augmente la taille de l'échantillon, plus la précision est grande (ou l'erreur d'estimation petite).



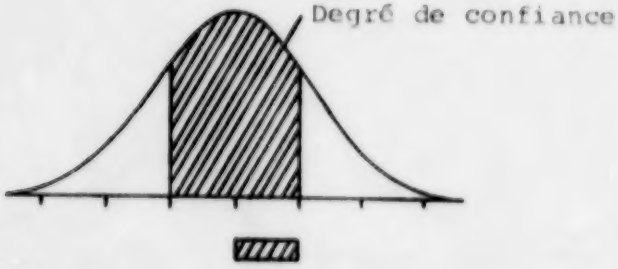


ainsi

$n_1 < n_2$ et $d_1 = d_2$ ⟹ ▨ < ▨

Alors pour une précision donnée et un degré de confiance souhaité, quelle doit être la taille de l'échantillon (n)?

n = ?

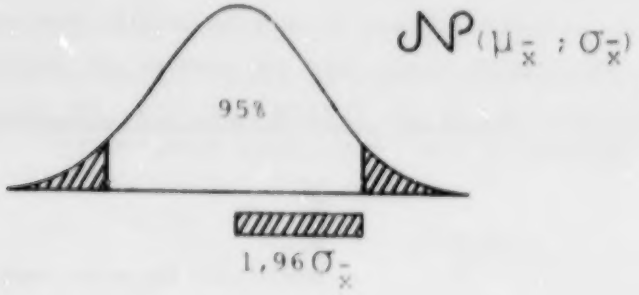


La précision (ou l'erreur acceptable) peut s'exprimer techniquement comme un nombre de fois l'écart-type de nos données.

$$\text{Erreur acceptable} = \text{précision} = z_c \sigma_{\bar{x}} = z_c \frac{\sigma}{\sqrt{n}}$$

(tirage avec remise)

Exemple: Si on veut avoir 95% de chance que l'erreur ne dépasse pas 2 unités, on peut trouver la taille de l'échantillon (n).



$$2 = 1,96\sigma_{\bar{x}} = 1,96 \frac{\sigma}{\sqrt{n}}$$

Ici, il peut arriver deux choses:

σ est connu ou il est inconnu.

1) Si σ est connu, disons $\sigma = 10$

$$2 = 1,96 \frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{10}{\sqrt{n}}$$

donc $n = 96,04$

(On choisit $n = 97$)

2) Si σ est inconnu, on se retrouve dans une impasse. Le seul moyen de s'en sortir est d'essayer d'estimer la valeur de σ à partir d'un échantillon. La dispersion de l'échantillon est presque toujours plus petite que la dispersion de la population. C'est pourquoi on a avantage à prendre comme estimé de σ une valeur un peu plus élevée que s.

En effet, s est la somme des carrés des écarts à la moyenne divisée par le nombre de données. Si au lieu de diviser par n on divise par n-1, on va obtenir une valeur plus élevée que s, donc plus près de σ.

Ainsi,

$$2 = 1{,}96 \frac{\sigma}{\sqrt{n}} - 1{,}96 \frac{\hat{\sigma}}{\sqrt{n}} \qquad (1)$$

où

$$\hat{\sigma} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \qquad (2)$$

Si le n obtenu par la formule (1) est plus grand que le n employé pour estimer σ avec $\hat{\sigma}$ (2), on doit augmenter la taille de notre échantillon pour obtenir le degré de confiance désiré sinon l'échantillon ayant servi à estimer σ est suffisant.

Résumé

n = ? l'erreur et le degré de confiance connus

1. μ = ?

$$\text{Erreur} = z_c \, \sigma_{\bar{x}}$$

i) z_c = ?

Si $n \geq 30$, Z suit une loi normale

$(z_c \sim Z)$

ii) $\sigma_{\bar{x}}$ = ?

a) Tirage avec remise

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad (r.1)$$

{1) σ est connu
{2) σ est inconnu

σ est estimé par $\hat{\sigma}$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

b) Tirage sans remise

On corrige la formule (r.1)

par $\sqrt{\frac{N-n}{N-1}}$

2. P = ?

$$\text{Erreur} = z_c \, \sigma_p$$

i) z_c = ?

Si $n \geq 30$, Z suit une loi normale

$(z_c \sim Z)$

Résumé (suite)

ii) σ_p = ?

a) Tirage avec remise

$$\sigma_p = \sqrt{\frac{PQ}{n}} \qquad (r.2)$$

1) P est connu, Q= 1-P

2) P est inconnu,

P est estimé par la valeur $\frac{1}{2}$.

(On obtient ainsi la plus grande valeur de n ce qui implique qu'on va obtenir le degré de confiance désiré.)

b) Tirage sans remise

On corrige la formule (r.2) par $\sqrt{\frac{N-n}{N-1}}$.

Comment estimer un paramètre d'une population?

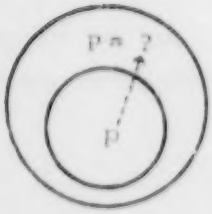
On peut faire deux genres d'estimation.

I. Une estimation ponctuelle

C'est-à-dire se servir de l'information obtenue de notre échantillon et l'inférer à la population.

Exemples:

- Se servir de la moyenne de l'échantillon comme estimé de la moyenne de la population.



- Se servir de la proportion de l'échantillon comme estimé de la proportion de la population.

- Se servir de la variance de l'échantillon comme estimé de la variance de la population.

On est conscient que cette façon de procéder implique des erreurs. Comme on ne peut l'empêcher, il faudrait éviter cependant des erreurs systématiques liées à cette façon de procéder.

Quelles sont les qualités d'un bon estimateur?

Elles se rapportent soit à la justesse de l'estimation obtenue, soit à sa précision.

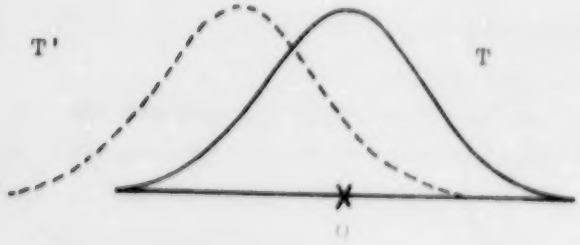
Exemple:

Une balance peut donner une évaluation correcte du poids (justesse) mais à un gramme près (précision).

Une autre balance peut fournir un résultat beaucoup plus précis à un centigramme près mais erroné.

1) L'estimateur est sans biais.

L'estimation obtenue est juste; sa valeur tourne autour de la vraie valeur du paramètre ( ).

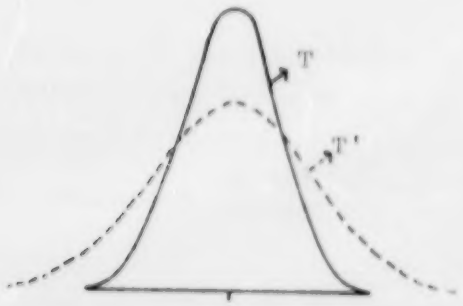


vraie valeur
du
paramètre

T: estimateur juste de
T': estimateur non correct de

2) L'estimateur est <u>efficace</u>.

Sa variance est petite. Plus l'estimateur est efficace, plus l'estimation est <u>précise</u>.



T est un estimateur plus efficace que T'.

En pratique on demande à un estimateur:

a) en premier lieu, d'être sans biais (justesse);

b) d'être aussi efficace que possible (précision).

- La moyenne d'échantillon ($\bar{x}$) remplit ces deux conditions pour estimer la moyenne de la population (μ).

- La variance d'échantillon (s^2) multipliée par $n/n-1$ remplit aussi ces deux conditions par rapport à la variance de la population (σ^2).

2. Une estimation d'intervalle de confiance

Principe: *On se sert de l'information fournie par l'échantillon.*

La valeur du paramètre peut être plus grande ou plus petite que la valeur fournie par l'échantillon; on lui ajoute et soustrait donc une certaine valeur pour obtenir le degré de confiance désiré.

Exemple: Estimation de μ à 95% de confiance.

1) Se servir de $\bar{x}$ comme indicateur.

2) Ajouter à $\bar{x}$ une valeur et lui soustraire la même valeur (v).

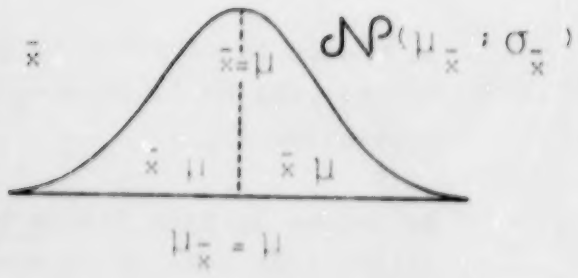
$$\underbrace{\bar{x} - v}_{L_{inf}} \qquad \underbrace{\bar{x} + v}_{L_{sup}}$$

On obtient ainsi deux limites:
une limite inférieure (L_{inf}) et
une limite supérieure (L_{sup}).

Le problème revient à chercher une valeur (v) telle que la moyenne de la population (μ) se trouve entre les deux limites (L_{inf} et L_{sup}) avec la probabilité de 95%.

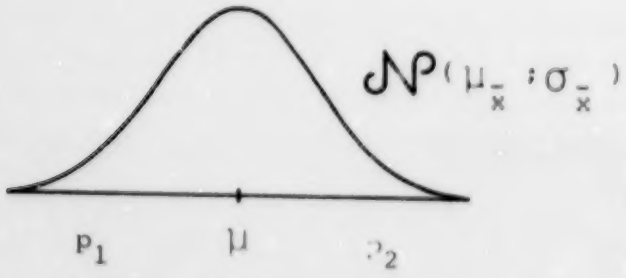
$$Pr\ (\ L_{inf} < \mu < L_{sup}\ ) = 95\%$$

Si la taille de l'échantillon (n) est égale ou supérieure à 30, on sait que les moyennes d'échantillons sont distribuées selon la loi normale autour de leur moyenne ($\mu_{\bar{x}}$) qui est égale à la moyenne de la population (μ).



En d'autres termes, il y a des échantillons dont leur moyenne est égale ($\bar{x} = \mu$) , d'autres supérieure ($\bar{x}$ μ) ou inférieure ($\bar{x}$ μ) à la moyenne de la population (μ).

L'échantillon qu'on a choisi au hasard est un de ces échantillons. On ignore cependant où se situe sa moyenne par rapport à ces moyennes d'échantillons. Ce qu'on sait, c'est la chance qu'elle soit à telle place.



- Sa chance la plus élevée est d'être située près de la moyenne de la population (μ).

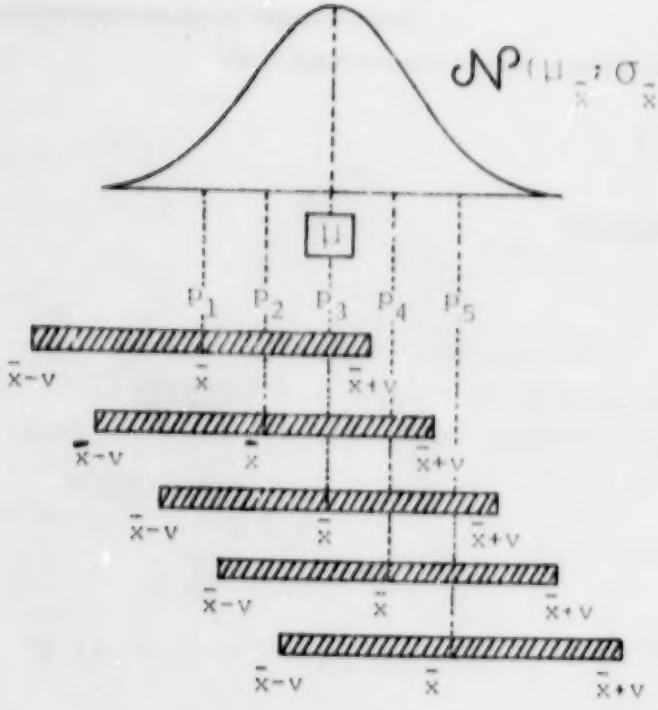
- Sa chance la plus faible est d'être située très loin de la moyenne de la population (μ) (à la position p_1 ou à la position p_2).

On veut dire que, peu importe l'endroit où la moyenne de l'échantillon ($\bar{x}$) se situe par rapport aux autres moyennes d'échantillons, on ait 95% de chance d'inclure la moyenne de la population (μ) dans les limites L_{inf} et L_{sup} lorsqu'on s'en sert comme indicateur.

Ainsi regardons visuellement ce qui se passe si on ajoute et retranche à la moyenne de l'échantillon ($\bar{x}$) une valeur (v) représentée ici par un bâton ▨.

Quelle que soit la position possible de la moyenne de l'échantillon ($\bar{x}$), la moyenne de la population (μ) est-elle toujours incluse dans les deux limites obtenues en ajoutant ou en soustrayant cette valeur (v) à la moyenne de l'échantillon ($\bar{x}$)?

a) Si la valeur (v) est ▨.

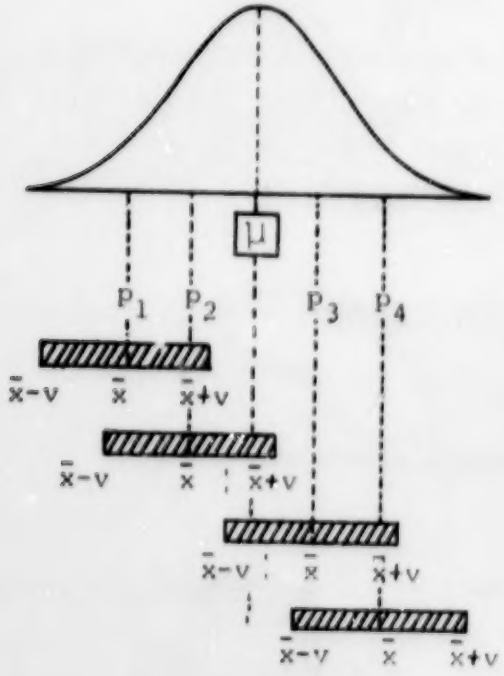


Position	Moyenne de la population incluse (μ)
P_1	Oui
P_2	Oui
P_3	Oui
P_4	Oui
P_5	Oui

Quel que soit l'endroit où peut se trouver la moyenne de l'échantillon ($\bar{x}$) quand on ajoute la valeur (v) (▨), on inclut toujours la moyenne de la population (μ) dans l'intervalle formé.

$$Pr(L_{inf} \leq \mu \leq L_{sup}) = 100\%$$

b) Si la valeur (v) est ▨



Position	Moyenne de la population incluse (μ)
P_1	non
P_2	oui
P_3	oui
P_4	non

Quel que soit l'endroit où peut se trouver la moyenne de l'échantillon ($\bar{x}$) quand on ajoute la valeur (v) (▨), on n'inclut pas toujours la moyenne de la population (μ) dans l'intervalle formé.

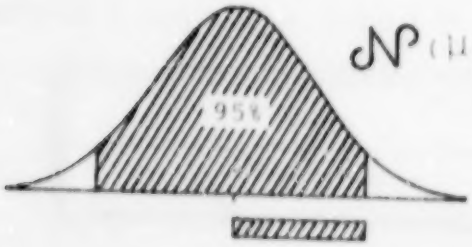
$$Pr(L_{inf} < \mu < L_{sup}) = 100\%$$

On a vu auparavant qu'on peut exprimer cette valeur (v) en termes de

$$z_c \sigma_{\bar{x}}$$

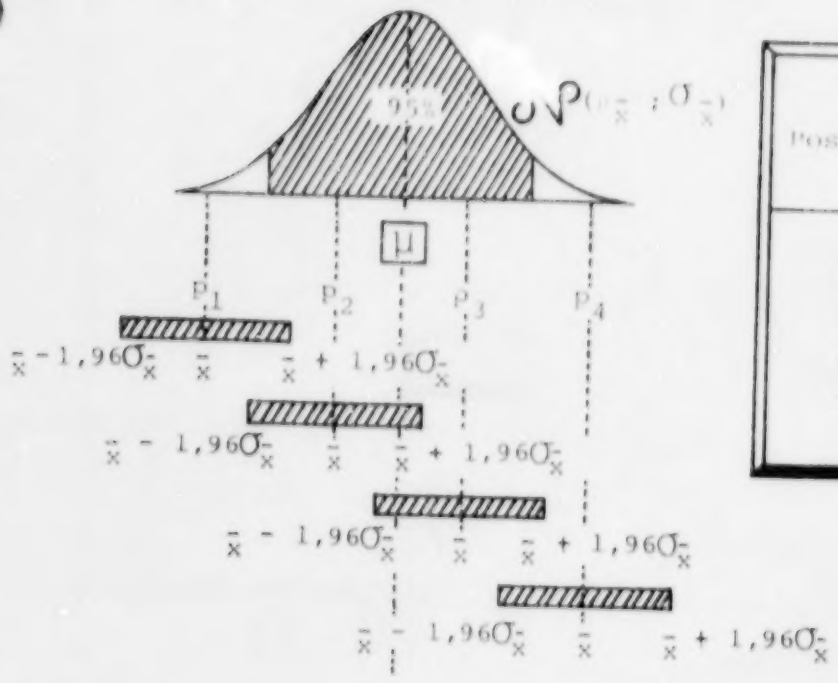
(elle correspond à l'erreur permise ou à la précision désirée).

Ainsi, si $z_c = 1,96$ et $v = 1,96\sigma_{\bar{x}}$ (▨).



$$Pr(\bar{x} - 1,96\sigma_{\bar{x}} \;\mu\; \bar{x} + 1,96\sigma_{\bar{x}}) = 95\%$$

En effet,



Position	Moyenne de la population incluse (μ)
p_1	non
p_2	oui
p_3	oui
p_4	non

La probabilité que la moyenne de l'échantillon ($\bar{x}$) soit autour de la position p_1 ou de la position p_4 (partie non hachurée) est de 5% et de 95% autour des autres positions comme p_2 et p_3 (partie hachurée).

Donc en ajoutant la valeur $1,96\sigma_{\bar{x}}$ à la moyenne de l'échantillon et en la soustrayant, on a 95% de chance que la moyenne de la population (μ) se trouve entre les 2 limites.

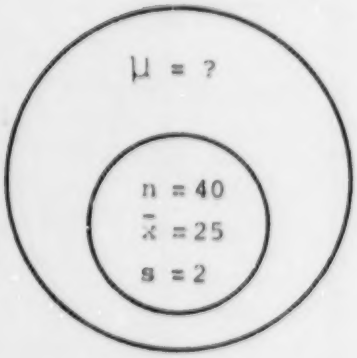
L_{inf}	$<$ μ $<$	L_{sup}
$\bar{x} - 1,96\sigma_{\bar{x}}$	$<$ μ $<$	$\bar{x} + 1,96\sigma_{\bar{x}}$

On a ainsi trouvé la valeur (v).

On peut faire de même pour tout autre degré de confiance et paramètre.

Exemple:

Estimation de μ à 99% à partir d'un échantillon (n= 40)



$\bar{x} - v$ $\quad$ $\bar{x} + v$

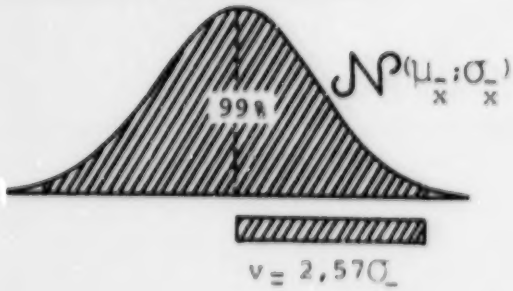
$25 - v$ $\quad$ $25 + v$

L_{inf} $\quad$ L_{sup}

$Pr(25-v \leq \mu \leq 25-v) = 99\%$

v = ?

$Pr(25-2,57\sigma_{\bar{x}} \leq \mu \leq 25-2,57\sigma_{\bar{x}}) = 99\%$



$25-2,57\frac{\sigma}{\sqrt{n}} \leq \mu \leq 25-2,57\frac{\sigma}{\sqrt{n}}$

(tirage avec remise)

$\hat{\sigma} = \sqrt{\frac{n}{n-1}}\, s = \sqrt{\frac{40}{39}} \times 2$

$25-2,57 \times \frac{2,02}{\sqrt{40}} \leq \mu \leq 25-2,57 \times \frac{2,02}{\sqrt{40}}$

$24,18 \leq \mu \leq 25,82$ à 99%

Résumé

ESTIMATION D'UN PARAMÈTRE DE LA POPULATION

1. μ = ?

   a) Estimation ponctuelle

      On utilise $\bar{x}$.

   b) Estimation par intervalle de confiance

      i) Se servir de $\bar{x}$.

      ii) Ajouter et soustraire à $\bar{x}$ une valeur selon le degré de confiance souhaité.

$$\bar{x} \pm z_c \sigma_{\bar{x}}$$

où $\sigma_{\bar{x}} = \begin{cases} \dfrac{\sigma}{\sqrt{n}} & \text{(tirage avec remise)} \\ \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}} & \text{(tirage sans remise)} \end{cases}$

2. σ = ?

   Estimation ponctuelle

   On utilise

$$\hat{\sigma} = s \sqrt{\frac{n}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}}$$

3. $P = ?$

a) Estimation ponctuelle

On se sert de p.

b) Estimation par intervalle de confiance

i) Se servir de p.

ii) Ajouter et soustraire à p une valeur selon le degré de confiance souhaité.

$$p \pm z_c \, \sigma_p$$

$$\sigma_p = \begin{cases} \sqrt{\frac{P\,Q}{n}} & \text{(tirage avec remise)} \\ \sqrt{\frac{P\,Q}{n}} \sqrt{\frac{N-n}{N-1}} & \text{(tirage sans remise)} \end{cases}$$

Comment tester une hypothèse statistique?

Une hypothèse statistique est un énoncé concernant le comportement d'une population et dont on se propose d'examiner la validité à partir de l'information obtenue d'un échantillon aléatoire pris dans cette population ou à partir des résultats d'un plan d'expérience.

Une hypothèse est paramétrique si elle porte sur la valeur d'un paramètre d'une variable de la population (comme l'âge moyen de tous les Québécois, la proportion de chômeurs au Québec, etc. ...) ou si elle porte sur la forme de distribution d'une variable de la population (par exemple, le quotient intellectuel des adultes suit une loi normale de moyenne 100 et d'écart-type 15,...).

Habituellement lors d'un test statistique, deux hypothèses sont confrontées: celle à laquelle on s'intéresse et son contraire. L'une d'elles est appelée hypothèse nulle (notée H_0) et l'autre contre-hypothèse (notée H_1).

Lorsque les hypothèses sont formulées on aura à les confronter avec les données de l'échantillon. Toutefois, nous sommes face à un problème important. Un échantillon ne reproduit pas nécessairement la population dont il provient, de façon exacte. Une certaine incertitude est associée au jugement que nous allons poser. Notre but est de quantifier le risque d'erreur.

Deux risques d'erreurs sont associés à la conclusion d'un test. L'erreur de première espèce qui consiste à rejeter l'hypothèse nulle alors qu'elle est vraie et l'erreur de seconde espèce qui consiste à rejeter la contre-hypothèse alors qu'elle est vraie.

La probabilité de faire une erreur de première espèce définit le niveau de signification du test (noté α : lire alpha). La probabilité de faire une erreur de seconde espèce dépend de la spécification de la contre-hypothèse.

On peut résumer les deux sortes d'erreurs dans le tableau suivant:

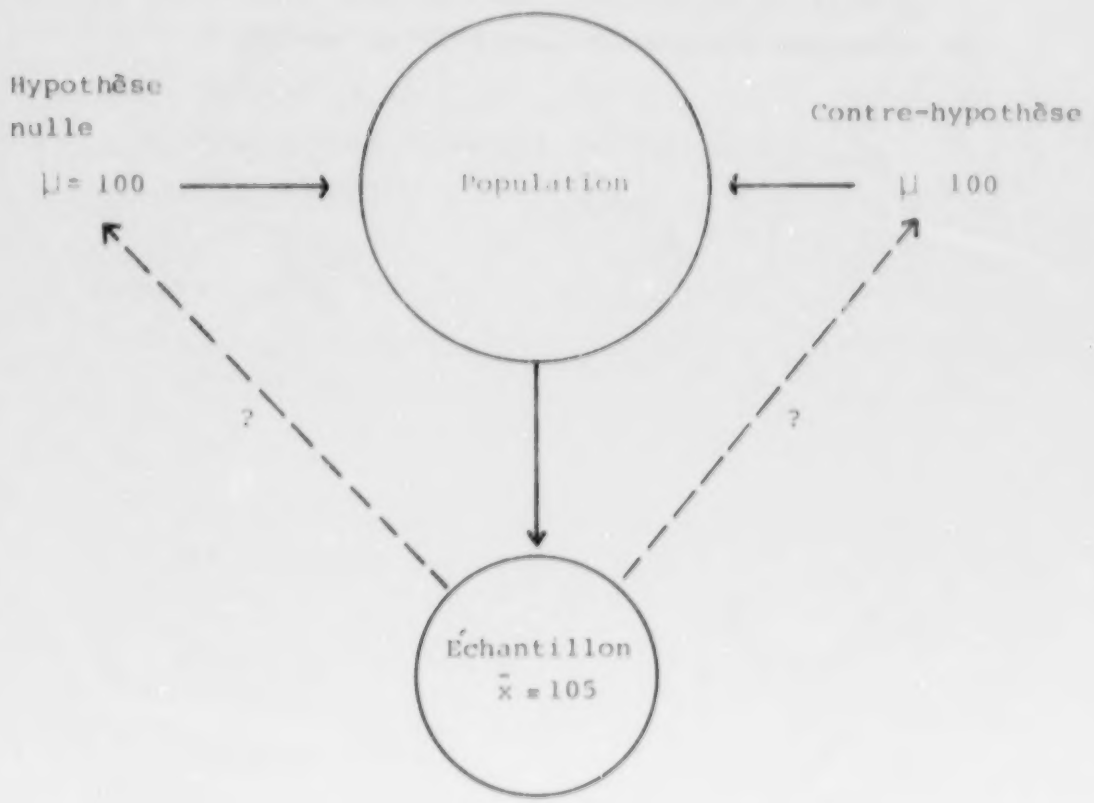
		Hypothèse vraie	
		Hypothèse nulle	Contre-hypothèse
Hypothèse retenue après le test	Hypothèse nulle	Aucune erreur	Erreur de seconde espèce
	Contre-hypothèse	Erreur de première espèce	Aucune erreur

Puisque le niveau de signification correspond à la probabilité de faire une erreur, on lui attribue une petite valeur (5%, 2%, 1%, ...). Si on lui donne la valeur 5%, ceci signifie qu'on a 5% des chances de rejeter l'hypothèse nulle alors qu'elle est vraie.

Habituellement, on essaie de minimiser autant que possible les deux risques d'erreur. Ce n'est pas simple car pour un échantillon d'une taille donnée, la décroissance d'un type d'erreur est en général accompagnée par la croissance de l'autre type d'erreur. La seule façon de diminuer les deux types d'erreur est d'augmenter la taille de l'échantillon et la seule façon de les éliminer est d'étudier la population complète, ce qui peut être long et onéreux.

Afin de prendre une décision sur l'acceptation d'une des deux hypothèses avancées, ça prend une règle. Un échantillon n'aura pas nécessairement la même moyenne que la population d'où il est prélevé. Il semble raisonnable d'accepter l'hypothèse nulle si la valeur de la moyenne de l'échantillon ne s'écarte pas trop de la valeur prévue par cette hypothèse. Par contre, si elle s'écarte trop de la valeur prévue et que ceci n'a pas beaucoup de chance de se produire, on dira que la différence entre la valeur obtenue et celle prévue est significative et on rejettera l'hypothèse nulle.

Par exemple:



La règle de décision définit deux régions complémentaires:

- La première, dite région d'acceptation de l'hypothèse nulle;

- la seconde, dite région de rejet de l'hypothèse nulle.

Ces régions dépendent entre autres du niveau de signification.

Selon l'énoncé de la contre-hypothèse, un test est unilatéral ou bilatéral. Dans un test bilatéral, la région de rejet est divisée en deux parties égales.

Étapes à suivre pour élaborer un test d'hypothèses

E.1 *Poser les hypothèses.*

E.2 *Choisir le niveau de signification du test.*

E.3 *Établir la règle de décision.*

Avant même de choisir l'échantillon, on doit établir la règle qui va nous permettre d'accepter l'une des deux hypothèses. Lors de cette étape on suppose vraie l'hypothèse nulle.

E.4 *Choisir un échantillon aléatoire dans la population et calculer la statistique appropriée: moyenne, proportion et autres.*

On doit fixer la taille de l'échantillon et s'assurer que l'échantillon est aléatoire.

E.5 *Prendre la décision.*

Nous allons illustrer les étapes à l'aide de deux exemples.

Exemple 1:

Dans une laiterie, une machine doit verser 1 litre de lait dans des contenants de carton. On veut vérifier si la machine est bien ajustée.

La population est constituée de tous les contenants remplis par la machine.

E.1 Poser les hypothèses

L'hypothèse nulle s'énonce ainsi:

En moyenne, l'ensemble des contenants remplis par la machine contient 1 litre et la machine est bien ajustée.

La contre-hypothèse est:

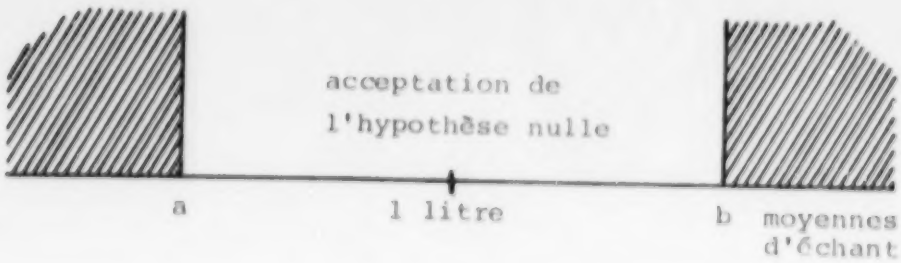
En moyenne, l'ensemble des contenants remplis par la machine ne contient pas 1 litre et la machine n'est pas bien ajustée.

E.2 Choisir le niveau de signification du test.

Fixons le niveau de signification du test à 5%.

E.3 Établir la règle de décision

Puisque les hypothèses portent sur la valeur de la moyenne d'une population, nous concentrerons notre attention sur la valeur de la moyenne de l'échantillon extrait, sans remise, de cette population. Si la moyenne de l'échantillon prend une valeur pas très différente de 1 litre, on va être porté à accepter l'hypothèse nulle. Dans le cas contraire, on sera porté à accepter la contre-hypothèse. Les régions d'acceptation et de rejet de l'hypothèse nulle peuvent être représentées ainsi:



La région hachurée est la région de rejet de l'hypothèse nulle. Ce test est bilatéral puisque la région de rejet est divisée en deux parties égales.

Il faut maintenant trouver les valeurs critiques a et b.

La théorie de l'échantillonnage des moyennes nous permet d'énoncer les faits suivants:

si $\bar{x}_i$ représente les moyennes des différents échantillons, de même taille, qu'on peut extraire, sans remise, de cette population,

alors

i) la moyenne des $\bar{x}_i$ égale la moyenne de la population (1 litre si H_0 est vraie) i.e.

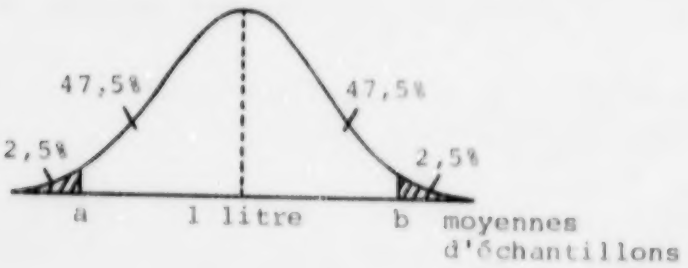
$$\mu_{\bar{x}} = 1 \text{ litre}$$

ii) l'écart-type des $\bar{x}_i$ vaut

$$\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}},$$

iii) les $\bar{x}_i$ suivent approximativement une courbe normale si $n \geq 30$, même si la forme de distribution de la population est inconnue.

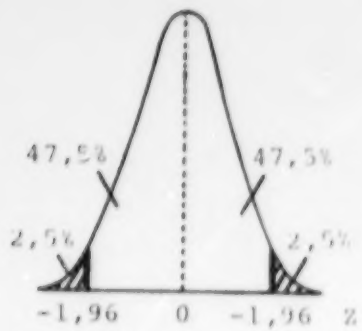
Graphiquement, on peut représenter ces faits de la façon suivante:



La surface de la région hachurée égale le niveau de signification (5% dans notre exemple). La surface de la région non hachurée vaut 95%.

L'écart-type de la population (σ) ici est inconnu. Toutefois, on a déjà vu qu'il peut être estimé par $\hat{\sigma}$ où $\hat{\sigma}$ est l'écart-type modifié de l'échantillon. La taille de la population (N) est inconnue. On peut affirmer que cette valeur est très grande. En effet, la machine doit remplir un très grand nombre de contenants. C'est pourquoi l'expression $\sqrt{\frac{N-n}{N-1}}$ vaut à peu près 1. Par conséquent l'écart-type des $\bar{x}_i$ vaut approximativement $\hat{\sigma} / \sqrt{n}$.

Si on convertit la courbe normale ci-dessus en courbe $\mathcal{N}_{(0;1)}$, les surfaces représentées sont égales.



On a les égalités suivantes:

$$\frac{a-1}{\hat{\sigma}/\sqrt{n}} = -1,96$$

$$\frac{b-1}{\hat{\sigma}/\sqrt{n}} = 1,96$$

D'où

$$a = 1-1,96\,\hat{\sigma}/\sqrt{n}$$
$$b = 1+1,96\,\hat{\sigma}/\sqrt{n}$$

La règle de décision s'énonce ainsi:

Si la moyenne de l'échantillon prend une valeur entre a et b, on va accepter l'hypothèse nulle; sinon on va accepter la contre-hypothèse.

E.4 Choisir un échantillon aléatoire dans la population et calculer la statistique appropriée: moyenne, proportion, et autres.

À l'étape précédente, on a constaté que la taille de l'échantillon (n) doit être supérieure ou égale à 30. Choisissons 36. Après avoir choisi l'échantillon, on mesure la quantité de lait (en litre) de chaque contenant. On calcule la moyenne ($\bar{x}$) et l'écart-type modifié ($\hat{\sigma}$) de ces mesures. Supposons que l'on obtient 1,002 litre comme moyenne et 0,012 litre comme écart-type modifié.

E.5 Prendre la décision.

Les valeurs de a et b sont:

$$a = 1 - \frac{1,96 \times 0,012}{\sqrt{36}} = 0,996 \text{ litre}$$

$$b = 1 + \frac{1,96 \times 0,012}{\sqrt{36}} = 1,004 \text{ litre}$$

Puisque la moyenne de l'échantillon prend une valeur (1,002 litre) entre 0,996 litre et 1,004 litre, on accepte l'hypothèse nulle i.e. la machine est bien ajustée, avec un seuil de signification de 5%.

Le fait d'accepter l'hypothèse nulle ne veut pas dire qu'elle est exacte, mais simplement qu'on estime ne pas avoir de raisons suffisantes de lui préférer la contre-hypothèse compte tenu des informations dont on dispose.

Exemple 2

On veut savoir si le taux de chômage au Québec est supérieur à 10%.

La population est composée de l'ensemble des Québécois(es) aptes à travailler (population active).

E.1 Poser les hypothèses.

L'hypothèse nulle est:

La proportion des chômeurs(euses) au Québec est égale à 10%.

La contre-hypothèse est:

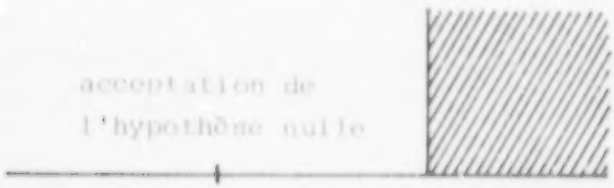
La proportion des chômeurs(euses) au Québec est supérieure à 10%.

E.2 Choisir le niveau de signification du test.

Fixons-le à 5%.

E.3 Établir la règle de décision.

Les hypothèses portent sur la proportion de chômeurs(euses) au Québec. Nous concentrerons notre attention sur la proportion de chômeurs dans un échantillon extrait, sans remise, de la population. Si cette dernière proportion prend une valeur près de 10%, on va être porté à accepter l'hypothèse nulle. Si elle est beaucoup supérieure à 10%, on sera porté à accepter la contre-hypothèse.



Le test est unilatéral à droite.

La théorie de l'échantillonnage des proportions nous permet d'énoncer les faits suivants:

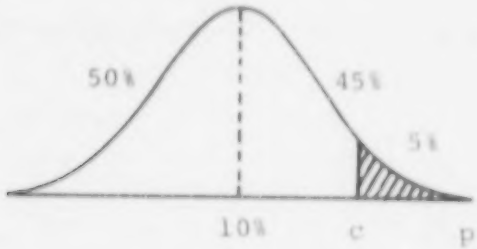
Les proportions (p) de chômeurs(euses) des différents échantillons, de même taille, qu'on peut extraire de la population

i) ont une moyenne (μ_p) égale à la proportion de chômeurs(euses) dans la population (10% si H_0 est vraie).

ii) ont un écart-type (σ_p) égal à $\sqrt{\frac{(0,1)(0,9)}{n}}\sqrt{\frac{N-n}{n-1}}$ si H_0 est vraie

iii) suivent approximativement une courbe normale si $n \geq 30$.

Graphiquement, on obtient:



Comme à l'exemple 1, puisque N est beaucoup plus grand que n, alors $\sqrt{\frac{N-n}{N-1}}$ vaut à peu près 1. Par conséquent,

$\sigma_p \approx \sqrt{\frac{(0,1)(0,9)}{n}}$.

Si on convertit la courbe normale ci-dessus en courbe $\mathcal{N}_{(0;1)}$, on a

$$\frac{c - 10\%}{\sqrt{\frac{0,09}{n}}} = 1,645$$

D'où $$c = 10\% + 1,645 \sqrt{\frac{0,09}{n}}$$

E.4 Choisir un échantillon aléatoire dans la population et calculer la statistique appropriée: moyenne, proportion et autres.

On doit choisir aléatoirement au moins trente Québécois(es) pour former l'échantillon. Choisissons-en 30 et calculons la proportion de chômeurs(euses) dans l'échantillon (p).

Supposons que cette proportion est 20%.

E.5 Prendre la décision.

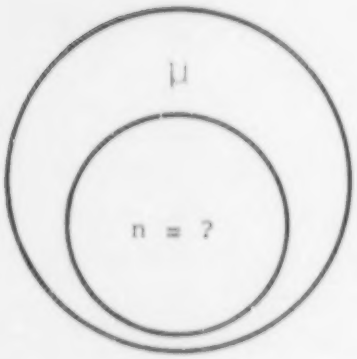
$$c = 10\% + 1{,}645\sqrt{\frac{0{,}09}{30}}$$

$$= 10\% + 9\% = 19\%$$

Puisque p, la proportion de chômeurs(euses) dans l'échantillon, vaut 20% et est supérieure à 19%, on rejette l'hypothèse nulle et on accepte la contre-hypothèse c'est-à-dire le taux de chômage au Québec est supérieur à 10% avec un seuil de signification de 5%.

Remarque: On peut accepter H_0 avec un seuil de signification et la refuser en changeant le seuil. Ainsi à l'exemple 2, si on fixe le seuil de signification à 1%, c vaudra $10\% + 2{,}33\sqrt{\frac{0{,}09}{30}}$ 22,8% et l'hypothèse nulle sera acceptée au seuil de signification de 1%.

Comment déterminer la taille d'un échantillon à partir de la connaissance des risques entre deux hypothèses simples?



H_o : $\mu = m_o$

H_1 : $\mu = m_1$

On veut déterminer à l'aide d'un échantillon laquelle des deux hypothèses est la plus vraisemblable.

On peut commettre deux genres d'erreur:

1) l'erreur de 1e espèce (α)

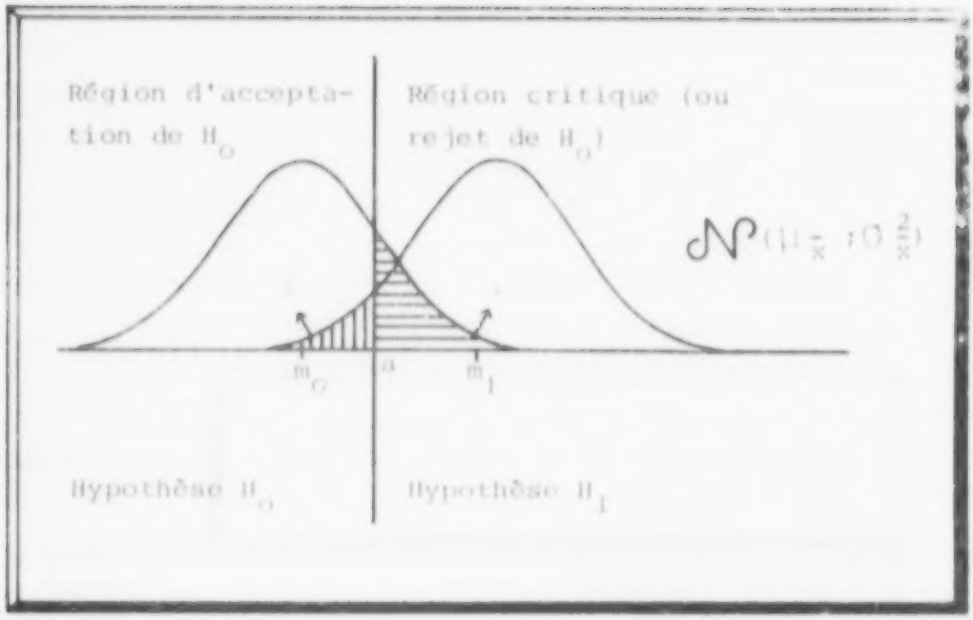
$$\alpha = \left[\Pr \quad \text{décider } H_1 / H_o \text{ vraie} \right]$$

2) l'erreur de 2e espèce (β)

$$\beta = \left[\Pr \quad \text{décider } H_o / H_1 \text{ vraie} \right]$$

Que doit être la taille de l'échantillon (n) pour avoir des erreurs de 1ère et 2ème espèces acceptables?

n = ?



Sous la forme standardisée,

$$\alpha = Pr\left[z > z_\alpha\right] \quad \text{avec} \quad z_\alpha = \frac{a-m_o}{\frac{\sigma}{\sqrt{n}}} \qquad (1)$$

$$\beta = Pr\left[z \leq z_\beta\right] \quad \text{avec} \quad z_\beta = \frac{a-m_1}{\frac{\sigma}{\sqrt{n}}} \qquad (2)$$

(avec remplacement)

Formons le rapport

$$\frac{z_\alpha}{z_\beta} = \frac{a-m_o}{a-m_1}$$

a = ?

$$z_\alpha \ (a-m_1) = z_\beta \ (a-m_o)$$

$$a\,z_\alpha - m_1 z_\alpha = a\,z_\beta - m_0\,z_\beta$$

$$a\,z_\alpha - a\,z_\beta = m_1\,z_\alpha - m_0\,z_\beta$$

Ainsi

$$a = \frac{m_1\,z_\alpha - m_0\,z_\beta}{z_\alpha - z_\beta}$$

En portant dans l'équation (1) ou l'équation (2), on obtient la taille n de l'échantillon.

$$n = \frac{\sigma^2\,z_\alpha^2}{(a - m_0)^2} = \frac{\sigma^2\,z_\alpha^2}{(a - m_1)^2}$$ (avec remplacement)

$$n = \frac{Nz_\alpha^2\,\sigma^2}{((a-m_0)^2(N-1) + (z_\alpha^2\,\sigma^2)} = \frac{Nz_\beta^2\,\sigma^2}{((a-m_0)^2(N-1) + z_\beta^2\,\sigma^2)}$$

(sans remplacement)

Exemple:

$n = ?$ $\quad \sigma^2 = 25$

$H_0 : \mu = 50 = m_0$

$H_1 : \mu = 52 = m_1$

$\alpha = 5\%$ $\quad \beta = 1\%$

$z_\alpha = 1{,}645$

$z_\beta = 2{,}33$

Il faut déterminer a

$$a = \frac{m_1 z_\alpha - m_0 z_\beta}{z_\alpha - z_\beta} = \frac{52 \times 1{,}645 - 50 \times (-2{,}33)}{1{,}645 - (-2{,}33)} = 50{,}82$$

D'où

$$n = \frac{\sigma^2\,z_\alpha^2}{(a-m_0)^2} = \frac{25 \times 1{,}645^2}{(50{,}82-50)^2} = 99{,}35 \qquad n = 100$$

2. La méthode pour étudier des données

Comment étudier statistiquement des données?

À partir du travail effectué sur l'enquête, on peut dégager une méthode pour étudier statistiquement des données. Cette méthode divisée en sept étapes peut être utilisée non seulement pour traiter une enquête mais aussi lors d'un contrôle de qualité d'un produit ou d'une expérimentation. Voici une description de chaque étape.

1. Définition de la problématique

Lors de cette étape, on doit bien identifier le sujet de l'étude, cerner et présenter le problème. On doit définir la population à laquelle l'étude s'adresse.

2. Description de la méthodologie

On doit déterminer la façon de résoudre le problème posé lors de l'étape précédente, en particulier comment recueillir l'information.

Cette information peut provenir de l'utilisation de documents existants (recensement gouvernemental, questionnaire déjà utilisé, ...), ou de la fabrication de documents inexistants (questionnaire, plan d'expérimentation, plan de contrôle de la qualité, ...).

Il faut décider si l'information sera recueillir auprès de toute la population ou d'un échantillon.

Dans ce dernier cas, on doit déterminer les éléments de la population qui vont faire partie de l'échantillon. On doit, si possible, trouver la taille de l'échantillon et évaluer le degré d'incertitude des résultats.

3. Cueillette des données

On recueille l'information selon la méthode explicitée à l'étape précédente. On peut faire passer un questionnaire, s'il s'agit d'une enquête, remplir une carte de contrôle dans le cas d'un contrôle de la qualité ou effectuer des tests s'il s'agit d'une expérimentation.

4. Dépouillement et codage des données

Afin de faciliter l'analyse des données qui est la prochaine étape, il est utile de résumer autant que possible, toute l'information dans un tableau d'observations. Souvent on doit alors transformer en codes numériques ou alphabétiques les différents éléments d'information obtenue. Ce travail est indispensable lorsqu'on veut traiter les données par informatique. Le tableau d'observations constitue le fichier à traiter.

5. Analyse statistique des données

L'analyse statistique des données comprend trois parties importantes: le plan d'analyse, le genre d'analyse et les techniques employées et enfin l'analyse des données proprement dite.

5.1 Plan d'analyse

On élabore le plan qui sera suivi lors de l'analyse des données. Quel aspect du problème sera étudié en premier, en deuxième, etc... Il est essentiel d'analyser un aspect du problème à la fois. Il ne faut pas oublier l'objectif poursuivi et les différentes questions à l'origine de la recherche.

5.2 Genre d'analyse et techniques employées

On doit décrire d'une façon succincte, les différents outils employés dans l'analyse.

5.3 Analyse des données proprement dite

On doit analyser l'information recueillie, les résultats obtenus à partir des outils employés en respectant le plan d'analyse établi.

6. Interprétation des résultats

Il faut revenir sur les résultats obtenus à l'analyse des données, les discuter, en ressortir les principaux éléments, les expliquer et les interpréter dans le contexte de l'étude.

7. Conclusion

Pour le sujet étudié, on doit identifier les principaux résultats obtenus selon la méthode employée.

3. La rédaction d'un rapport

Comment rédiger un rapport?

La valeur d'une étude statistique repose en bonne partie sur la qualité technique de la présentation des résultats obtenus.

Ce texte présente quelques grandes lignes et techniques à respecter. Pour plus de détails, on peut toujours se référer à des ouvrages spécialisés.

BERNIER, B. (1967). Présentation d'un travail de recherche. Université Laval (Bibliothèque générale).

LORTIE, M. (1964). Le rapport technique et scientifique. Les Presses de l'Université Laval.

A) Les grandes lignes du rapport

1. L'avant-propos ou la préface

Cette partie est facultative et elle ne doit pas être confondue avec l'introduction; elle est essentiellement un discours préliminaire placé en tête du rapport et dont le rôle n'est pas d'amorcer le sujet à traiter. L'auteur y mentionne les raisons qui l'ont amené à étudier le sujet choisi, le but qu'il poursuit, ainsi que l'ampleur et les limites de son travail. Il peut y ajouter les remerciements qu'il juge bon d'adresser.
Si l'auteur tient la préface pour superflu, il peut quand même remercier qui de droit, mais dans ce cas, la page doit être intitulée "REMERCIEMENTS".

2. L'introduction

L'auteur présente le sujet d'une façon générale, définit la problématique et les objectifs de l'étude.

3. Le corps du rapport

Chapitre I - LES DONNÉES ET LA MÉTHODOLOGIE

L'auteur décrit la méthodologie, la cueillette des données, le dépouillement et le codage des données s'il y a lieu.

Chapitre II- L'ANALYSE STATISTIQUE DES DONNÉES

L'auteur présente le plan d'analyse, le genre d'analyse, les techniques employées et l'analyse des données proprement dites. Les calculs doivent être annexés.

Chapitre III- DISCUSSION DES RÉSULTATS

L'auteur interprète les résultats de l'étude.

4. Conclusion

L'auteur rappelle le sujet étudié, la méthode employée et les principaux résultats.

5. Les annexes

L'auteur peut y inclure des matières qui ne sont pas strictement nécessaires à la compréhension du texte, des textes complémentaires, des graphiques ou des tableaux dont le format dépasse les dimensions du texte et les calculs.

B) Les principales techniques

1. La pagination

On ne pagine jamais les pages qui portent un titre (page titre, pages initiales de l'avant-propos, de la table des matières, de l'introduction, de chacun des chapitres, etc.), mais elles sont toutefois comptées dans la pagination.

Le folio prend normalement place en haut de page, à un demi-pouce de son bord supérieur, soit au centre de la page, soit dans le coin supérieur droit à un demi-pouce environ du bord. Les pages qui portent un titre font cependant exception à cette règle; le folio se place dans ce cas en bas au centre de la page, à moins qu'on ne préfère le supprimer totalement.

On pagine en chiffres romains minuscules toute la partie préliminaire et en chiffres arabes tout le reste du rapport à partir de l'introduction jusqu'à la fin, les annexes également. Les pages de

tableaux, graphiques insérés dans le texte doivent être paginées à la suite comme les pages de texte afin de pouvoir y référer dans le texte.

2. Les nombres

Dans un texte où les nombres sont rares, on se sert du langage des mots pour exprimer les nombres inférieurs à la centaine et de celui des chiffres arabes pour les autres, mais dans un contexte où figurent des nombres inférieurs et supérieurs à cent, on n'utilise partout que des chiffres. On ne commence jamais une phrase par un ou plusieurs chiffres; on reconstruit donc la phrase pour éviter pareil début.

3. Les références

Les coutumes relatives aux références bibliographiques diffèrent quelque peu d'une discipline à l'autre.

Dans le domaine des sciences expérimentales:

a) Les noms d'auteurs mentionnés dans le texte sont toujours suivis de l'année de publication de l'ouvrage ou article auquel on réfère. (Exemple: "Briand (1982) a montré que..."). Un numéro entre parenthèses réfère au numéro correspondant de la "Liste des ouvrages cités" placés à la fin.

b) Dans la "Bibliographie", les noms d'auteurs sont placés, si possible, par ordre alphabétique des

noms d'auteurs, suivis dans l'ordre de l'année de publication, du titre de l'ouvrage ou de l'article (souligné), du nom de la revue, du volume, du numéro (entre parenthèses), et des pages (exemple: "PROULX, S.M. (1982). Une nouvelle synthèse de l'acide aminoadipique. Canadian Journal of Research, Section B, 27 (1): 21-24").

Si l'ouvrage a été édité, on mentionne, à la suite du titre, le nom et l'adresse de l'éditeur (exemple: "Masson, Paris").

4. La mise en page

Les chapitres, divisions et subdivisions de chapitre doivent être clairement indiquées en utilisant soit le système classique de division (I, II,... A, B,... 1, 2, ... a, b, ...) soit le système décimal (1, 2,... 1.1, 1.2, 1.2.1, ...). Chaque chapitre commence sur une nouvelle page.

Le rapport doit être présenté dans l'ordre suivant:

a) Une feuille de garde, blanche
b) Une page titre
c) L'avant-propos ou la préface, s'il y a lieu
d) La table des matières
e) La liste des tableaux, des figures

(Les pages portant la matière qui précède (de b à f inclusivement) sont paginées au moyen de chiffres romains minuscules, tandis que toutes les pages subséquentes sont numérotées au moyen de chiffres arabes.)

f) L'introduction
g) Le corps de la thèse
h) La conclusion
i) La bibliographie ou la liste des ouvrages cités
j) Les annexes

C) Exemple

(Une feuille de garde, blanche)

(MODÈLE DE PAGE TITRE)

CÉGEP DE RIMOUSKI

RAPPORT

PRÉSENTÉ À

JOSEPH-MARIE BRIAND

COMME EXIGENCE PARTIELLE

AU COURS DE STATISTIQUE

PAR

(Nom de l'étudiant/e)

LES ACTIVITÉS ÉTUDIANTES DU GROUPE AU COURS DES DERNIÈRES VACANCES ESTIVALES

MAI 1982

AVANT-PROPOS

(MODÈLE DE TABLE DES MATIÈRES)

TABLE DES MATIÈRES

PAGE

AVANT-PROPOS ii

TABLE DES MATIÈRES iii

LISTE DES TABLEAUX iv

LISTE DES FIGURES v

INTRODUCTION 1

Chapitre

1. LES DONNÉES ET LA MÉTHODOLOGIE 3

1.1 Les données 3

1.2 La méthodologie 5

2. L'ANALYSE STATISTIQUE DES DONNÉES 7

3. LA DISCUSSION DES RÉSULTATS 15

CONCLUSION 18

BIBLIOGRAPHIE 20

ANNEXES 22

1 GRAPHIQUES DES REVENUS HEBDOMADAIRES 25

2 PROGRAMME TRAIT 30

(MODÈLE DE LISTE DES TABLEAUX)

LISTE DES TABLEAUX

No		Page
1	Le pourcentage d'étudiants(es) qui ont exercé les activités: travail, études, voyages et autres	8
2	Le revenu gagné par les étudiants(es) pendant leurs dernières vacances	10
	etc.	

(MODÈLE DE LISTE DES FIGURES)

LISTE DES FIGURES

No		Page
1	Le pourcentage d'étudiants(es) qui ont exercé les activités: travail, études, voyages	9

etc.

BIBLIOGRAPHIE

CAILLEZ, F., PAGES S.P. (1976). Introduction à l'analyse des données. SMASH.

CLÉROUX, R. (1976). Analyse de données unidimensionnelles provenant d'une expérience planifiée. Notes de cours IFT 6660, Université de Montréal.

EL-SABH, M.I., LÉVESQUE, L., MURTY, T.S. (1979). Mouvements des eaux induits par la marée et le vent dans l'estuaire du Saint-Laurent. Naturaliste can., 106 (1-2): 89-104.

MURTY, T.S., EL-SABH, M.I., BRIAND, J.M. (1979). Storm surge amplitudes in the St-Lawrence Estuary. Lighthouse, 20: 8-10.

etc.

SYMBOLES

A_0^z	Aire sous la courbe normale entre 0 et z.
α (alpha)	Probabilité de l'erreur de première espèce, niveau de signification.
β (bêta)	Probabilité de l'erreur de deuxième espèce.
c_i	Valeur de la ième catégorie.
Cov	Covariance.
e_i	Effectif de la ième catégorie.
E_i	Événement.
E.M.	Écart moyen.
f_i	Fréquence.
H_0	Hypothèse nulle.
H_1	Hypothèse contraire (rivale).
m_i	Point milieu d'une classe.
md	Médiane.
mo	Mode.
nc	Nombre de catégories.

ncl	Nombre de classes.
μ (mu)	Moyenne arithmétique de la population.
μ_p	Moyenne des proportions d'échantillons.
$\mu_{\bar{x}}$	Moyenne de la distribution d'échantillonnage des moyennes.
n	Taille (effectif) de l'échantillon.
N	Taille de la population.
P	Probabilité de succès. Proportion de réussite d'un événement dans un échantillon.
par	Paramètre.
Pr(a < x < b)	Probabilité que x soit compris entre a et b.
$Pr(E_i)$	Probabilité de l'événement E_i.
$Pr(E_i/E_j)$	Probabilité de l'événement E_i étant donné la réussite de E_j.
p	Proportion d'un événement dans la population.
Q_i	Ième quantile.
ϕ	Événement impossible.
r	Coefficient de corrélation linéaire.
R	Étendue simple.

s	Écart-type d'un échantillon.
s^2	Variance d'un échantillon.
S	Ensemble fondamental ou événement certain.
σ (sigma)	Écart-type de la population.
σ^2	Variance de la population.
$\sigma_{\bar{x}}$	Écart-type de la distribution des moyennes d'échantillons.
σ_p	Écart-type des proportions d'échantillons
$\hat{\sigma}$	Estimation de l'écart-type de la population avec un échantillon.
$\sigma_{\bar{x}}^2$	Variance de la distribution des moyennes d'échantillons.
σ_p^2	Variance de la distribution des proportions d'échantillons.
Σ	Sommation.
v	Valeur de la précision ou de l'erreur acceptable.
$\bar{x}$	Moyenne arithmétique de l'échantillon.
z_c	Valeur critique de z qui détermine le degré de confiance
z	Variable standardisée.